

工藤
04
932

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 3 月 3 1 日
Date of Application:

出 願 番 号 特 願 2 0 0 3 - 0 9 4 2 8 1
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 0 9 4 2 8 1]

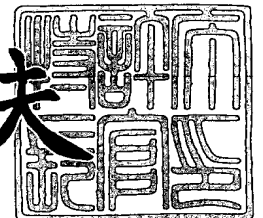
出 願 人 日 本 電 気 株 式 有 限 公 司
Applicant(s):



2 0 0 3 年 1 0 月 2 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 3 - 3 0 8 1 3 1 9

【書類名】 特許願

【整理番号】 65700149

【提出日】 平成15年 3月31日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/00

【発明者】

 【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

 【氏名】 藤田 悟

【特許出願人】

 【識別番号】 000004237

 【氏名又は名称】 日本電気株式会社

【代理人】

 【識別番号】 100102864

 【弁理士】

 【氏名又は名称】 工藤 実

【手数料の表示】

 【予納台帳番号】 053213

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9715177

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 コンピュータシステム、コンピュータプログラム、コンピュータ間の通信方法、構造化文書の符号化方法、符号化された構造化文書の復号方法

【特許請求の範囲】

【請求項 1】 構造化文書の内部表現を符号化して符号化文書を作成する符号化手段を含む送信側コンピュータと、

ネットワークを介して前記符号化文書を取得し、前記符号化文書を復号して前記内部表現を再構成する復号手段を含む受信側コンピュータとを備え、

前記符号化手段は、前記構造化文書に含まれる記述を重複することなく列挙することによって前記内部表現から検証支援情報を生成する検証支援情報抽出手段を含み、且つ、前記検証支援情報を含むように前記符号化文書を生成し、

前記復号手段は、前記検証支援情報に基づいて、再構成される前記内部表現のデータ構造が、所定の文書型宣言に適合しているかを検証する検証手段を含むコンピュータシステム。

【請求項 2】 請求項 1 に記載のコンピュータシステムにおいて、

前記検証支援情報は、前記内部表現を構成する要素のうち、テキストデータを要素内容として含むテキスト含有要素の要素テキストが、重複しないように列挙された要素テキストリストを含み、

前記文書型宣言は、前記テキスト含有要素の型を定義した要素型宣言を含み、

前記検証手段は、前記要素テキストが、前記型に適合するか否かを検証するコンピュータシステム。

【請求項 3】 請求項 1 に記載のコンピュータシステムにおいて、

前記検証支援情報は、前記内部表現を構成する要素のうち、属性を有する属性所有要素と、前記属性の属性値である属性テキストとが、重複しないように列挙された属性テキストリストを含み、

前記文書型宣言は、前記属性の属性型を定義した属性型宣言を含み、

前記検証手段は、前記属性テキストが、前記属性型に適合するか否かを検証する

コンピュータシステム。

【請求項 4】 請求項 1 に記載のコンピュータシステムにおいて、
前記検証支援情報は、前記内部表現に含まれる部分構造が、重複しないように
列挙された部分構造リストを含み、

前記検証手段は、前記部分構造が、前記文書型宣言に定義された構文に適合す
るか否かを検証する

コンピュータシステム。

【請求項 5】 請求項 4 に記載のコンピュータシステムにおいて、
前記検証支援情報は、
前記要素の要素名を列挙した要素名リストと、
前記属性の属性名を列挙した属性名リストと、
前記内部表現を構成する要素のうち、テキストデータを要素内容として含むテ
キスト含有要素の要素テキストが、重複しないように列挙された要素テキス
トリストと、

前記内部表現を構成する要素のうち、属性を有する属性所有要素と、前記属性
の属性値である属性テキストとが、重複しないように列挙された属性テキス
トリスト

とを含み、

前記部分構造のそれぞれには、構造 ID が与えられ、

前記要素テキストと前記属性テキストとのそれぞれにテキスト ID が与えられ

、
前記部分構造の部分構造表現は、前記要素の要素 ID と、前記要素が有する属
性の属性 ID と、前記要素の子要素の要素 ID と、前記親要素が、要素テキ
ストを含むことを示す記号とによって記述され、

前記符号化文書は、前記構造 ID と前記テキスト ID とによって前記内部表現
の全体の構造を表現する全体構造データを含み、

前記復号手段は、前記要素名リストと前記属性名リストと前記全体構造デー
タに基づいて前記内部表現を再構成する

コンピュータシステム。

【請求項 6】 構造化文書の内部表現を符号化して符号化文書を作成する符号化手段を備え、

前記符号化手段は、前記構造化文書に含まれる記述を重複することなく列挙することによって前記内部表現から検証支援情報を生成する検証支援情報抽出手段を含み、且つ、前記検証支援情報を含むように前記符号化文書を生成するコンピュータ。

【請求項 7】 請求項 6 に記載のコンピュータにおいて、

前記検証支援情報は、前記内部表現を構成する要素のうち、テキストデータを要素内容として含むテキスト含有要素の要素テキストが、重複しないように列挙された要素テキストリストを含むコンピュータ。

【請求項 8】 請求項 6 に記載のコンピュータにおいて、

前記検証支援情報は、前記内部表現を構成する要素のうち、属性を有する属性所有要素と、前記属性の属性値である属性テキストとが、重複しないように列挙された属性テキストリストを含むコンピュータ。

【請求項 9】 請求項 6 に記載のコンピュータにおいて、

前記検証支援情報は、前記内部表現に含まれる部分構造が、重複しないように列挙された部分構造リストを含むコンピュータ。

【請求項 10】 請求項 9 に記載のコンピュータにおいて、

前記検証支援情報は、
前記要素の要素名を列挙した要素名リストと、
前記属性の属性名を列挙した属性名リストと、
前記内部表現を構成する要素のうち、テキストデータを要素内容として含むテキスト含有要素の要素テキストが、重複しないように列挙された要素テキストリストと、

前記内部表現を構成する要素のうち、属性を有する属性所有要素と、前記属性の属性値である属性テキストとが、重複しないように列挙された属性テキストリ

スト

とを含み、

前記部分構造のそれぞれには、構造 I D が与えられ、

前記要素テキストと前記属性テキストとのそれぞれにテキスト I D が与えられ

、

前記部分構造の部分構造表現は、前記要素の要素 I D と、前記要素が有する属性の属性 I D と、前記要素の子要素の要素 I D と、前記親要素が、要素テキストを含むことを示す記号とによって記述され、

前記符号化文書は、前記構造 I D と前記テキスト I D とによって前記内部表現の全体の構造を表現する全体構造データを含む

コンピュータ。

【請求項 11】 ネットワークを介して符号化文書を取得し、前記符号化文書を復号して構造化文書の内部表現を再構成する復号手段を備え、

前記符号化文書は、前記構造化文書に含まれる記述が重複することなく列挙された検証支援情報を含み、

前記復号手段は、前記検証支援情報に基づいて、再構成される前記内部表現のデータ構造が、所定の文書型宣言に適合しているかを検証する検証手段を含むコンピュータ。

【請求項 12】 請求項 11 に記載のコンピュータにおいて、

前記検証支援情報は、前記内部表現を構成する要素のうち、テキストデータを要素内容として含むテキスト含有要素の要素テキストが、重複しないように列挙された要素テキストリストを含み、

前記文書型宣言は、前記テキスト含有要素の型を定義した要素型宣言を含み、

前記検証手段は、前記要素テキストが、前記型に適合するか否かを検証するコンピュータ。

【請求項 13】 請求項 11 に記載のコンピュータにおいて、

前記検証支援情報は、前記内部表現を構成する要素のうち、属性を有する属性所有要素と、前記属性の属性値である属性テキストとが、重複しないように列挙

された属性テキストリストを含み、

前記文書型宣言は、前記属性の属性型を定義した属性型宣言を含み、

前記検証手段は、前記属性テキストが、前記属性型に適合するか否かを検証する

コンピュータ。

【請求項 14】 請求項 11 に記載のコンピュータにおいて、

前記検証支援情報は、前記内部表現に含まれる部分構造が、重複しないように
列挙された部分構造リストを含み、

前記検証手段は、前記部分構造が、前記文書型宣言に定義された構文に適合する
か否かを検証する

コンピュータ。

【請求項 15】 請求項 14 に記載のコンピュータにおいて、

前記検証支援情報は、

前記要素の要素名を列挙した要素名リストと、

前記属性の属性名を列挙した属性名リストと、

前記内部表現を構成する要素のうち、テキストデータを要素内容として含むテキスト含有要素の要素テキストが、重複しないように列挙された要素テキストリストと、

前記内部表現を構成する要素のうち、属性を有する属性所有要素と、前記属性の属性値である属性テキストとが、重複しないように列挙された属性テキストリスト

とを含み、

前記部分構造のそれぞれには、構造 ID が与えられ、

前記要素テキストと前記属性テキストとのそれぞれにテキスト ID が与えられ

、

前記部分構造の部分構造表現は、前記要素の要素 ID と、前記要素が有する属性の属性 ID と、前記要素の子要素の要素 ID と、前記親要素が、要素テキストを含むことを示す記号とによって記述され、

前記符号化文書は、前記構造 ID と前記テキスト ID とによって前記内部表現

の全体の構造を表現する全体構造データを含み、

前記復号手段は、前記要素名リストと前記属性名リストと前記全体構造データとに基づいて前記内部表現を再構成する

コンピュータ。

【請求項 16】 (a) 構造化文書の内部表現を符号化して符号化文書を作成するステップと、

(b) ネットワークを介して前記符号化文書を取得し、前記符号化文書を復号して前記内部表現を再構成するステップ

とを備え、

前記 (a) ステップは、

(a1) 前記構造化文書に含まれる記述を重複することなく列挙することによって前記内部表現から検証支援情報を生成するステップと、

(a2) 前記検証支援情報を含むように前記符号化文書を生成するステップとを含み、

前記 (b) ステップは、

(b1) 前記検証支援情報に基づいて、再構成される前記内部表現のデータ構造が、所定の文書型宣言に適合しているかを検証するステップ

を含む

構造化文書通信方法。

【請求項 17】 (a1) 構造化文書に含まれる記述を重複することなく列挙することによって前記構造化文書の内部表現から検証支援情報を生成するステップと、

(a2) 前記検証支援情報を含むように符号化文書を生成するステップとをコンピュータに実行させる

コンピュータプログラム。

【請求項 18】 (b) ネットワークを介して符号化文書を取得し、符号化文書を復号して構造化文書の内部表現を再構成するステップ

をコンピュータに実行させるコンピュータプログラムであって、

前記符号化文書は、前記構造化文書に含まれる記述が重複することなく列挙さ

れた検証支援情報を含み、

前記 (b) ステップは、

(b 1) 前記検証支援情報に基づいて、再構成される前記内部表現のデータ構造が、所定の文書型宣言に適合しているかを検証するステップ

を含む

コンピュータプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、コンピュータシステムに関し、特に、構造化文書（構造化テキスト）を相互に通信するコンピュータで形成されたコンピュータシステムに関する。

【0002】

【従来の技術】

構造化文書（構造化テキスト）は、文書のデータ構造をあらわす記述（タグ）を文書自体に埋め込んだ文書である。構造化文書は、データ構造について高い柔軟性と拡張性とを有しており、このような特長から、構造化文書は、異なるコンピュータ、及び異なるアプリケーション間のデータ交換に広く使用される。XML、及びSGMLは、典型的な構造化文書の規格である。

【0003】

構造化文書の構造を規定するために、文書型定義が使用される。例えば、XML文書の文書型定義としてDTD (Document Type Definition) 及びXMLスキーマが使用される。文書型定義は、構造化文書自体に埋め込まれることも、構造化文書とは別に用意されることもある。構造化文書とは別途に用意されて文書型定義を記述したファイルは、文書構造定義ファイルと呼ばれる。

【0004】

構造化文書は、文書型定義に即した形式で正しく記述される必要がある。構造化文書が文書型定義に従わずに記述されていることは、その構造化文書を取り扱うコンピュータが該構造化文書の内容を正しく把握することを妨げる。このため、構造化文書は、そのデータ構造の妥当性が検証されることが望ましい。特許文

献1は、構造化文書のデータ構造の検証について開示している。

【0005】

構造化文書の検証における一つの問題は、構造化文書を検証するために必要な演算量が多いことである。構造化文書は、本質的に冗長性が大きく、その大きさが大きい傾向にある。この構造化文書の冗長性は、検証に必要な演算量を増加させ、このため、検証に必要なコストを増大させる。

【0006】

このような問題は、特に、構造化文書を一のコンピュータから他のコンピュータに伝送する場合に顕著になる。構造化文書を伝送する場合、該構造化文書は、送信側コンピュータと受信側コンピュータの両方で検証され得る。送信側コンピュータにおいて構造化文書の検証が行われても、受信した構造化文書のデータの完全性を保証するためには、受信側コンピュータにおいても構造化文書の検証が行われる必要がある。送信側コンピュータと受信側コンピュータの両方で構造化文書の検証が行われることは、検証に必要な演算量が多いという問題を更に深刻にする。

【0007】

【特許文献1】

特開 2001-75958号公報

【0008】

【発明が解決しようとする課題】

本発明の目的は、構造化文書を検証するために必要な演算量を抑制し、これにより検証に必要なコストを減少することにある。

本発明の他の目的は、構造化文書を送信側コンピュータから受信側コンピュータに伝送する場合に、受信側コンピュータが構造化文書を検証するために必要な演算量を抑制し、これにより検証に必要なコストを減少することにある。

【0009】

【課題を解決するための手段】

以下に、[発明の実施の形態]で使用される番号・符号を用いて、課題を解決するための手段を説明する。これらの番号・符号は、[特許請求の範囲]の記載

と「[発明の実施の形態]」の記載との対応関係を明らかにするために付加されている。但し、付加された番号・符号は、「[特許請求の範囲]」に記載されている発明の技術的範囲の解釈に用いてはならない。

【0010】

本発明によるコンピュータシステムは、構造化文書の内部表現（13）を符号化して符号化文書（4）を作成する符号化手段（12）を含む送信側コンピュータ（1）と、ネットワーク（3）を介して符号化文書（4）を取得し、符号化文書（4）を復号して内部表現（23）を再構成する復号手段（22）を含む受信側コンピュータ（2）とを備えている。符号化手段（12）は、該構造化文書に含まれる記述を重複することなく列挙することによって内部表現（13）から検証支援情報（4a～4e）を生成する検証支援情報抽出手段（15）を含み、且つ、検証支援情報（4a～4e）を含むように符号化文書（4）を生成する。復号手段（22）は、検証支援情報（4a～4e）に基づいて、再構成される内部表現（23）のデータ構造が、所定の文書型宣言に適合しているかを検証する検証手段（25）を含む。当該コンピュータシステムでは、記述の重複が排除された検証支援情報（4a～4e）に基づいて内部表現（23）のデータ構造の妥当性が検証される。このような構成は、検証手順の重複を排除し、従って、構造化文書を検証するために必要な演算量を抑制することができる。

【0011】

検証支援情報（4a～4e）は、内部表現（13）を構成する要素のうち、テキストデータを要素内容として含むテキスト含有要素の要素テキストが、重複しないように列挙された要素テキストリスト（4c）を含み、文書型宣言は、テキスト含有要素の型を定義した要素型宣言を含み、検証手段（25）は、該要素テキストが、該型に適合するか否かを検証することが好適である。これにより、同一の要素テキストが内部表現（13）に複数含まれていても、該要素テキストの検証の重複を避けることができる。

【0012】

検証支援情報（4a～4e）は、内部表現（13）を構成する要素のうち、属性を有する属性所有要素と、該属性の属性値である属性テキストとが、重複しな

いように列挙された属性テキストリスト (4 d) を含み、文書型宣言は、属性の属性型を定義した属性型宣言を含み、検証手段 (2 5) は、属性テキストが、属性型に適合するか否かを検証することが好適である。これにより、同一の属性テキストが内部表現 (1 3) に複数含まれていても、該属性テキストの検証の重複を避けることができる。

【0013】

検証支援情報 (4 a ~ 4 e) は、内部表現 (1 3) に含まれる部分構造が、重複しないように列挙された部分構造リスト (4 e) を含み、検証手段 (2 5) は、該部分構造が、文書型宣言に定義された構文に適合するか否かを検証することが好適である。これにより、同一の部分構造が内部表現 (1 3) に複数含まれていても、該部分構造の検証の重複を避けることができる。

【0014】

検証支援情報 (4 a ~ 4 e) が、該要素の要素名を列挙した要素名リスト (4 a) と、該属性の属性名を列挙した属性名リスト (4 b) と、要素テキストリスト (4 c) と、属性テキストリスト (4 d) とを含む場合、該部分構造のそれぞれには、構造 ID が与えられ、該要素テキストと該属性テキストとのそれぞれにテキスト ID が与えられ、該部分構造の部分構造表現は、該要素の要素 ID と、該要素が有する属性の属性 ID と、該要素の子要素の要素 ID と、該親要素が、要素テキストを含むことを示す記号とによって記述され、符号化文書 (4) は、該構造 ID と該テキスト ID とによって内部表現 (1 3) の全体の構造を表現する全体構造データ (4 f) を含み、復号手段 (2 2) は、要素名リスト (4 a) と属性名リスト (4 b) と全体構造データ (4 f) とに基づいて内部表現 (2 3) を再構成することが好適である。このような構成は、内部表現 (1 3、2 3) の全体構成を、小さなデータ量で表現することを可能にする。

【0015】

本発明によるコンピュータは、構造化文書の内部表現を符号化して符号化文書 (4) を作成する符号化手段 (1 2) を備えている。符号化手段 (1 2) は、該構造化文書に含まれる記述を重複することなく列挙することによって内部表現から検証支援情報 (4 a ~ 4 e) を生成する検証支援情報抽出手段 (2 5) を含み

、且つ、検証支援情報（4 a～4 e）を含むように符号化文書（4）を生成する。このようなコンピュータは、上記のコンピュータシステムを好適に実現する。

【0016】

本発明によるコンピュータは、ネットワーク（3）を介して符号化文書（4）を取得し、符号化文書（4）を復号して構造化文書の内部表現を再構成する復号手段（22）を備えている。符号化文書（4）は、該構造化文書に含まれる記述が重複することなく列挙された検証支援情報（4 a～4 e）を含んでいる。

復号手段（22）は、検証支援情報（4 a～4 e）に基づいて、再構成される内部表現（23）のデータ構造が、所定の文書型宣言に適合しているかを検証する検証手段（25）を含む。このようなコンピュータは、上記のコンピュータシステムを好適に実現する。

【0017】

本発明による構造化文書通信方法は、

（a）構造化文書の内部表現（13）を符号化して符号化文書（4）を作成するステップと、

（b）ネットワーク（3）を介して符号化文書（4）を取得し、符号化文書（4）を復号して内部表現（23）を再構成するステップとを備えている。

該（a）ステップは、

（a1）前記構造化文書に含まれる記述を重複することなく列挙することによって内部表現（13）から検証支援情報（4 a～4 e）を生成するステップと、

（a2）前記検証支援情報（4 a～4 e）を含むように前記符号化文書（4）を生成するステップ

とを含む。該（b）ステップは、

（b1）検証支援情報（4 a～4 e）に基づいて、再構成され内部表現（23）のデータ構造が、所定の文書型宣言に適合しているかを検証するステップを含む。このような構造化文書通信方法は、記述の重複が排除された検証支援情報（4 a～4 e）に基づいて内部表現（23）のデータ構造の妥当性を検証する。従って、当該構造化文書通信方法は、検証手順の重複を排除し、従って、構造

化文書を検証するために必要な演算量を抑制することができる。

【0018】

本発明によるコンピュータプログラムは、

(a1) 構造化文書に含まれる記述を重複することなく列挙することによって該構造化文書の内部表現(13)から検証支援情報(4a~4e)を生成するステップと、

(a2) 検証支援情報(4a~4e)を含むように符号化文書(4)を生成するステップ

とをコンピュータ(1)に実行させる。当該コンピュータプログラムは、上述のコンピュータシステムの実現に好適に使用される。

【0019】

本発明によるコンピュータプログラムは、

(b) ネットワーク(3)を介して符号化文書(4)を取得し、符号化文書(4)を復号して構造化文書の内部表現(23)を再構成するステップをコンピュータに実行させるコンピュータプログラムである。符号化文書(4)は、該構造化文書に含まれる記述が重複することなく列挙された検証支援情報(4a~4e)を含む。該(b)ステップは、

(b1) 検証支援情報(4a~4e)に基づいて、再構成される内部表現(23)のデータ構造が、所定の文書型宣言に適合しているかを検証するステップを含む。当該コンピュータプログラムは、上述のコンピュータシステムの実現に好適に使用される。

【0020】

【発明の実施の形態】

以下、添付図面を参照しながら、本発明の実施の一形態を説明する。本発明によるコンピュータシステムの実施の一形態では、図1に示されているように、送信側コンピュータ1と受信側コンピュータ2とが設けられる。

【0021】

送信側コンピュータ1には、構造化文書、(例えば、XML文書、及びSGML文書)に対応したアプリケーションプログラム11と構造化文書符号化モジュ

ール 12 とがインストールされている。アプリケーションプログラム 11 は、受信側コンピュータ 2 に送信しようとする構造化文書の内部表現である構造化文書内部表現 13 を生成する。図 10 に示されているように、アプリケーションプログラム 11 は、典型的には、外部から構造化文書 13' を受け取り、受け取った構造化文書 13' の構造化文書内部表現 13 を生成するプログラムである。

【0022】

図 1 に示されているように、送信側コンピュータ 1 には、送信側コンピュータ 2 が取り扱う構造化文書の構造を示す文書型宣言を記述した文書構造定義ファイル 14 が用意されることがある。文書構造定義ファイル 14 に記述される文書型宣言には、構造化文書内部表現 13 が表現する構造化文書の文書構造の規則が記述される。より具体的には、文書型宣言は、構造化文書に含まれる要素の型を定義する要素型宣言、要素に与えられる属性の型を定義する属性型宣言を含んでいる。

【0023】

文書構造定義ファイル 14 が用意される場合、構造化文書内部表現 13 は、文書構造定義ファイル 14 に記述されている文書型宣言に従って記述される必要がある。当該コンピュータシステムが XML を使用する場合、文書型宣言としては、例えば、DTD、XML スキーマが使用される。

【0024】

構造化文書符号化モジュール 12 は、アプリケーションプログラム 11 から渡された構造化文書内部表現 13 を符号化して符号化文書 4 を生成するコンピュータプログラムである。生成された符号化文書 4 は、ネットワーク 3 を介して受信側コンピュータ 2 に送信される。

【0025】

受信側コンピュータ 2 には、アプリケーションプログラム 21 と構造化文書復号化モジュール 22 とがインストールされている。構造化文書復号化モジュール 22 は、受信側コンピュータ 2 が受信した符号化文書 4 を復号化して、構造化文書内部表現 23 を生成するために使用される。アプリケーションプログラム 21 は、復号化文章復号化モジュール 21 から構造化文書内部表現 23 を受け取り、

所望の処理を行う。図10に示されているように、アプリケーションプログラム21は、典型的には、構造化文書内部表現23から、それによって表現されている構造化文書24'を再生するプログラムである。送信側コンピュータ1から受信側コンピュータ2への通信の間にエラーが発生しなければ、構造化文書内部表現23は、送信側コンピュータ1のアプリケーションプログラム11が生成する構造化文書内部表現13に一致する。

【0026】

図1に示されているように、受信側コンピュータ2は、受信側コンピュータ2が取り扱う構造化文書が従うべき文書型宣言を記述した文書構造定義ファイル24を保存している。送信側コンピュータ1に用意され得る文書構造定義ファイル14と同様に、文書構造定義ファイル24に記述されている文書型宣言は、構造化文書内部表現23に表現された構造化文書の文書構造の規則を記述している。より具体的には、文書構造定義ファイル24に記述されている文書型宣言は、構造化文書に含まれる要素の型を定義する要素型宣言、要素に与えられる属性の型を定義する属性型宣言を含んでいる。送信側コンピュータ1に文書構造定義ファイル14が用意される場合、受信側コンピュータ2に用意される文書構造定義ファイル24と送信側コンピュータ1に用意される文書構造定義ファイル14とは、一致される必要がある。

【0027】

受信側コンピュータ2が構造化文書内部表現23の内容を正しく把握するためには、構造化文書内部表現23が文書構造定義ファイル24に含まれる文書型宣言に従って記述されていること（このような構造化文書内部表現23の性質は、「構造化文書内部表現23の妥当性」と呼ばれる。）が保証されなくてはならない。構造化文書内部表現23がXML文書の内部表現である場合、これは、構造化文書内部表現23によって表現されるXML文書が「妥当なXML文書（valid XML document）」であることを保証することに相当する。

【0028】

しかし、既述のように、構造化文書は冗長であるから、構造化文書内部表現23を直接に検証して構造化文書内部表現23の妥当性を保証することは、それに

必要な演算量が大きく好ましくない。

【0029】

構造化文書内部表現 23 の妥当性を保証するために、符号化文書 4 が下記に述べられるような特別な形式で生成され、その符号化文書 4 に基づいて構造化文書内部表現 23 の妥当性が検証される。必要な演算量を抑えて構造化文書内部表現 23 の妥当性の保証を容易化するために、符号化文書 4 は、下記の 2 つの特性を備えている。第 1 に、符号化文書 4 は、検証支援情報を含んで構成されている。検証支援情報とは、受信側コンピュータ 2 において構造化文書内部表現 23 の妥当性の保証に使用される情報である。第 2 に、符号化文書 4 は、構造化文書内部表現 13 の冗長性を排除しながら、構造化文書内部表現 13 と同等の内容を記述している。既述のように、構造化文書の冗長性は、その検証に必要な演算量を増加させる。冗長性が少ない符号化文書 4 に基づいて構造化文書内部表現 23 の妥当性を保証することにより、妥当性の保証に必要な演算量が抑制される。

【0030】

図 2 は、符号化文書 4 の具体的な構造を示している。符号化文書 4 は、要素名リスト 4 a と、属性名リスト 4 b と、要素テキストリスト 4 c と、属性テキストリスト 4 d と、部分構造リスト 4 e と、全体文書構造データ 4 f とを含んでいる。

【0031】

要素名リスト 4 a は、送信側コンピュータ 1 のアプリケーションプログラム 11 によって提供される構造化文書内部表現 13 に含まれる要素の要素名を列举したリストである。図 3 は、要素名リスト 4 a の内容の例を示している。要素名リスト 4 a は、構造化文書内部表現 13 に含まれる要素のそれぞれについて、要素に固有に与えられた要素 ID と、その要素の要素名とを記述している。図 3 を参照すると、例えば、要素 ID "E01" が与えられた要素の要素名は、"要素名 A" である。

【0032】

要素名リスト 4 a において、同一の要素名は、重複して列举されない。ある要素名を有する要素が構造化文書内部表現 13 に複数回現れても、要素名リスト 4

a には、その要素名は一度しか現れない。

【 0 0 3 3 】

属性名リスト 4 b は、構造化文書内部表現 1 3 中で使用されている属性の属性名を列挙したリストである。図 4 は、属性名リスト 4 b の内容の例を示している。属性名リスト 4 a は、構造化文書内部表現 1 3 中で使用されている属性のそれぞれに固有に与えられた属性 I D と、その属性の属性名とを記述している。図 4 を参照すると、例えば、属性 I D ” A 0 1 ” が与えられている属性の属性名は、” 属性名 A ” である。

【 0 0 3 4 】

属性名リスト 4 b において、同一の属性名は、重複して列挙されない。ある属性名を有する属性が構造化文書内部表現 1 3 に複数回現れても、属性名リスト 4 b には、その属性名は一度しか現れない。

【 0 0 3 5 】

要素テキストリスト 4 c は、構造化文書内部表現 1 3 に含まれる要素の要素テキストを列挙したリストである。要素テキストとは、テキストデータである要素内容である。図 5 は、要素テキストリスト 4 c の内容の例を示している。要素テキストリスト 4 c は、構造化文書内部表現 1 3 に含まれる全要素のうち、要素テキストを含む要素の要素 I D と、その要素に含まれる要素テキストと、その要素テキストに固有に与えられたテキスト I D とを記述している。構造化文書では、ある要素の要素内容は、子要素のみで構成されることが許されているから、全ての要素が要素テキストリスト 4 c に記載されているとは限らないことに留意されたい。図 5 を参照すると、例えば、要素 I D ” E 0 1 ” が与えられている要素の要素内容は、テキスト I D ” T 0 1 ” が与えられている要素テキスト” テキスト A ” と、テキスト I D ” T 0 2 ” が与えられている要素テキスト” テキスト B ” と、テキスト I D ” T 0 3 ” が与えられている要素テキスト” テキスト C ” とを含んでいる。

【 0 0 3 6 】

要素テキストリスト 4 c において、同一の要素テキストは、重複して列挙されない。ある要素テキストが構造化文書内部表現 1 3 に複数回現れても、要素テキ

ストリスト 4 c には、その要素テキストは一度しか現れない。

【0037】

属性テキストリスト 4 d は、構造化文書内部表現 13 中で使用されている属性の属性テキストを列挙したリストである。ある属性の属性テキストとは、その属性の属性値を示すテキストデータである。図 6 は、属性テキストリスト 4 d の内容の例を示している。属性テキストリスト 4 d は、構造化文書内部表現 13 中で使用されている属性の属性 ID と、その属性の属性テキストと、その属性テキストに固有に与えられたテキスト ID とを記述している。ある属性テキストのテキスト ID は、他の属性テキストのテキスト ID と異なるとともに、上述の要素テキストのテキスト ID と異なるように定められる。

【0038】

図 6 を参照すると、例えば属性 ID "A 0 1" が与えられている属性の属性テキスト（属性値）は、テキスト ID "T 0 1" が与えられている属性テキスト "テキスト A" である。

【0039】

属性テキストリスト 4 d において、同一の属性テキストは、重複して列挙されない。ある属性テキストが構造化文書内部表現 13 に複数回現れても、属性テキストリスト 4 c には、その属性テキストは一度しか現れない。

【0040】

部分構造リスト 4 e は、構造化文書内部表現 13 に現れる部分構造を示す部分構造表現を列挙したリストである。部分構造は、一の要素（親要素）と、その要素の子要素とから構成される。部分構造には、親要素の子要素の子要素は、含まれないことに留意されたい。部分構造表現は、このような部分構造を表現する記載である。ある要素（親要素）の部分構造表現は、

- (1) 親要素の要素 ID
- (2) 親要素に規定されている属性の属性 ID、
- (2) 親要素に含まれる子要素の要素 ID
- (3) 親要素が要素テキストを含む場合、該親要素が要素テキストを含むことを示す記号

を含む。

【0041】

図7は、部分構造リスト4eの内容の例を示す。部分構造リスト4eは、部分構造のそれぞれに固有に与えられた構造IDとその部分構造の部分構造表現とを記述している。部分構造表現の先頭には、その部分構造を構成する親要素の要素IDが記述される。図7を参照すると、例えば構造ID”S01”が与えられている部分構造は、要素ID”E01”を有する親要素を含む。その親要素には、属性ID”A01”を有する属性が規定され、更に、その親要素は、属性ID”E02”を有する子要素を2つ含んでいる。また、構造ID”S04”が与えられている部分構造は、要素ID”E01”を有する親要素を含む。この要素は、ある要素テキストを含む。図7において記号”T”は、要素ID”E01”を有する要素が、要素テキストを含むことを意味している。構造ID”S04”が与えられている部分構造表現には、親要素以外の要素IDは含まれておらず、これは、該親要素は子要素を含まないことを示している。

【0042】

部分構造リスト4eにおいて、同一の部分構造は、重複して列挙されない。ある部分構造が構造化文書内部表現13に複数回現れても、部分構造リスト4eには、その部分構造は一度しか現れない。

【0043】

全体文書構造データ4fは、構造化文書内部表現13の全体の構造を示すデータであり、構造IDとテキストIDとによって記述される。全体文書構造データ4fには、構造IDに続いて、該構造IDで特定される部分構造の親要素の要素テキスト、又は、その親要素に与えられている属性の属性テキスト（属性値）のテキストIDが記述される。例えば、構造ID”S04”によって特定される部分構造の部分構造表現が、”E04 T”であり、要素ID”E04”によって特定される要素名が”D”であり、テキストID”T04”によって特定される要素テキストが、”テキストD”であるとする。全体文書構造データ4fが”S04 T04”という記載を含む場合、その記載は、要素名が”D”である要素の要素テキストが”テキストD”であることが表している。構造IDとテキスト

IDとによって記述される全体文書構造データ4 fは、より小さなデータ量で、構造化文書内部表現13の全体の構造を記述することができる。

【0044】

上述の要素名リスト4 aと、属性名リスト4 bと、要素テキストリスト4 cと、属性テキストリスト4 dと、部分構造リスト4 eと、全体文書構造データ4 fとは、符号化データ4を復号化して構造化文書内部表現13と同一の構造化文書内部表現23を復元するために必要な情報を全て含んでいる。

【0045】

要素名リスト4 aと、属性名リスト4 bと、要素テキストリスト4 cと、属性テキストリスト4 dと、部分構造リスト4 eとは、構造化文書内部表現23を復元するのに必要な情報であるとともに、受信側コンピュータ2における構造化文書内部表現23の妥当性の保証に使用される検証支援情報としても機能する。要素名リスト4 aと要素テキストリスト4 cとは、構造化文書内部表現23に含まれている要素の要素テキストが、その要素について文書構造定義ファイル24に宣言されている要素型に適合することを検証するために使用される。属性名リスト4 bと属性テキストリスト4 dとは、構造化文書内部表現23で使用されている属性の属性テキストが、その属性について文書構造定義ファイル24に宣言されている属性型に適合することを検証するために使用される。部分構造リスト4 eは、構造化文書内部表現23で使用されている部分構造が、文書構造定義ファイル24に定義されている文書構造に適合することを検証するために使用される。これらを検証することにより、符号化文書4を復号化して構造化文書内部表現23を復元する前に、構造化文書内部表現23の妥当性を保証することが可能になる。

【0046】

上述のように、要素テキストリスト4 c、属性テキストリスト4 d、及び部分構造リスト4 eからは、重複が排除されている。これは、構造化文書内部表現23の妥当性を保証するために必要な演算の演算量を有効に減少する。要素テキストリスト4 cにおいて、同一の要素テキストが重複して列挙されないことは、該要素テキストの妥当性の検証を重複して行う必要をなくす。更に、属性テキスト

リスト 4 d において、同一の属性テキストが重複して列挙されないことは、該属性テキストの妥当性の検証を重複して行う必要をなくす。加えて、部分構造リスト 4 e において、同一の部分構造が重複して列挙されないことは、該部分構造の妥当性の検証を重複して行う必要をなくす。このように、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e における重複の排除により、構造化文書内部表現 23 の妥当性を保証するために必要な演算の演算量が抑制される。

【0047】

図 8 は、符号化文書 4 を生成する構造化文書符号化モジュール 12 の構成を示す。構造化文書符号化モジュール 12 は、検証支援情報抽出モジュール 15 と、全体文書構造抽出モジュール 16 と、符号化文書出力モジュール 17 とを備えている。

【0048】

検証支援情報抽出モジュール 15 は、要素名／属性名抽出モジュール 15 a と、テキスト抽出モジュール 15 b と、部分構造抽出モジュール 15 c とを含む。要素名／属性名抽出モジュール 15 a は、構造化文書内部表現 13 から、要素名と属性名とを抽出して、要素名リスト 4 a と属性名リスト 4 b とを作成する。テキスト抽出モジュール 15 b は、構造化文書内部表現 13 から要素テキスト及び属性テキストとを抽出して、要素テキストリスト 4 c と属性テキストリスト 4 d とを作成する。要素テキストリスト 4 c と属性テキストリスト 4 d との作成には、要素名／属性名抽出モジュール 15 a によって生成された要素名リスト 4 a と属性名リスト 4 b が使用される。部分構造抽出モジュール 15 c は、構造化文書内部表現 13 から部分構造を抽出して部分構造リスト 4 e を作成する。部分構造抽出モジュール 15 c は、抽出した部分構造のそれぞれに構造 ID を与え、更に、部分構造の構造 ID とその部分構造表現とを列挙して部分構造リスト 4 e を作成する。

【0049】

要素名／属性名抽出モジュール 15 a、テキスト抽出モジュール 15 b、及び部分構造抽出モジュール 15 c は、送信側コンピュータ 1 が文書構造定義ファイ

ル 14 を有している場合に文書構造定義ファイル 14 を利用して要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e を効率的に作成するように構成されていることが好適である。

【0050】

全体文書構造抽出モジュール 16 は、構造化文書内部表現 13 から全体構造データ 4 f を作成する。全体文書構造抽出モジュール 16 は、構造化文書内部表現 13 に基づいて構造化文書の全体構造を把握し、その全体構造を構造 ID とテキスト ID とで表現する全体構造データ 4 f を生成する。

【0051】

符号化文書出力モジュール 17 は、要素名リスト 4 a と、属性名リスト 4 b と、要素テキストリスト 4 c と、属性テキストリスト 4 d と、部分構造リスト 4 e と、全体構造データ 4 f とを統合して、符号化文書 4 を生成する。

【0052】

図 9 は、符号化文書 4 を復号して構造化文書内部表現 23 を再生する構造化文書復号化モジュール 22 の構成を示す。構造化文書復号化モジュール 22 は、検証モジュール 25 と、符号化文書入力モジュール 26 と、構造化文書内部表現出力モジュール 27 とを備えている。

【0053】

符号化文書入力モジュール 26 は、符号化文書 4 を展開して、要素名リスト 4 a と、属性名リスト 4 b と、要素テキストリスト 4 c と、属性テキストリスト 4 d と、部分構造リスト 4 e と、全体構造データ 4 f とを生成する。

【0054】

検証モジュール 25 は、テキスト型検証モジュール 25 a と部分構造検証モジュール 25 b とを含む。テキスト型検証モジュール 25 a は、要素テキストリスト 4 c に列挙された要素テキストと、属性テキストリスト 4 d に列挙された属性テキストとが、文書構造定義ファイル 24 に記述されている文書型宣言に適合しているかを検証する。部分構造検証モジュール 25 b は、部分構造リスト 4 e に列挙された部分構造が、文書構造定義ファイル 24 に規定されている型に適合し

ているかを検証する。

【0055】

構造化文書内部表現出力モジュール 27 は、要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、部分構造リスト 4 e 及び全体構造データ 4 f から構造化文書内部表現 23 を再生する。

【0056】

続いて、本実施の形態におけるコンピュータシステムの動作が説明される。図 11 及び図 12 は、構造化文書符号化モジュール 12 が構造化文書内部表現 13 から符号化文書 4 を生成するために行う処理を示すフローチャートである。構造化文書符号化モジュール 12 に構造化文書内部表現 13 が与えられ、構造化文書符号化モジュール 12 の実行が開始されると（ステップ S01）、構造化文書符号化モジュール 12 は、何も列挙されていない要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e を生成する。更に、構造化文書符号化モジュール 12 は、構造化文書内部表現 13 に記述されている要素のうち、要素名／属性名抽出モジュール 15 a が処理すべき要素を選択する。選択された要素は、以下において、処理対象要素と呼ばれる。構造化文書符号化モジュール 12 の実行が開始された直後においては、ルート要素（即ち、最上位の要素）が処理対象要素として選択される。

【0057】

要素名／属性名抽出モジュール 15 a は、処理対象要素の要素名を抽出する（ステップ S02）。更に、要素名／属性名抽出モジュール 15 a は、処理対象要素の要素名が要素名リスト 4 a に登録されているかを調べる（ステップ S03）。登録されていない場合、要素名／属性名抽出モジュール 15 a は、処理対象要素の要素名を、処理対象要素の要素 ID とともに要素名リスト 4 a に登録する（ステップ S04）。ステップ S03、S04 の処理により、要素名は、重複することなく要素名リスト 4 a に登録される。

【0058】

更に、要素名／属性名抽出モジュール 15 a は、処理対象要素が属性を持つかを調べる（ステップ S05）。処理対象要素が属性を持たない場合には、構造化

文書符号化モジュール 12 の処理は、図 12 に示されているステップ S 14 にジャンプする。

【0059】

処理対象要素が属性を持つ場合には、要素名／属性名抽出モジュール 15 a は、処理対象要素が有する属性の一つを選択する（ステップ S 06）。処理対象要素が属性を一つしか有しない場合には、その属性が自動的に選択される。

【0060】

続いて、要素名／属性名抽出モジュール 15 a は、選択された属性の属性名を、構造化文書内部表現 13 から抽出する（ステップ S 07）。更に、要素名／属性名抽出モジュール 15 a は、選択された属性の属性名が属性名リスト 4 b に登録されているかを調べる。（ステップ S 08）。登録されていない場合、要素名／属性名抽出モジュール 15 a は、選択された属性の属性名を、その属性の属性 ID とともに属性名リスト 4 b に登録する（ステップ S 09）ステップ S 08、S 09 の処理により、属性名は、重複することなく属性名リスト 4 b に登録される。

【0061】

更に、要素名／属性名抽出モジュール 15 a は、処理対象要素に規定されている全ての属性についてステップ S 07 からステップ S 12 の処理、即ち、属性登録処理が行われたかを調べる。属性登録処理が全ての属性について行われていない場合、属性登録処理が行われていない属性について属性登録処理が行われる。

【0062】

処理対象要素に規定されている全ての属性について属性登録処理が行われると、図 12 に示されているように、テキスト抽出モジュール 15 b は、処理対象要素が要素テキストを有するかを調べる（ステップ S 14）。有しない場合、構造化文書符号化モジュール 12 の処理は、ステップ S 18 にジャンプする。処理対象要素が要素テキストを有する場合、テキスト抽出モジュール 15 b は、その要素テキストが要素テキストリスト 4 c に登録されているかを調べる（ステップ S 11）。登録されていない場合、テキスト抽出モジュール 15 b は、その要素テキストにテキスト ID を与え、更に、その要素テキストを、処理対象要素の要素

IDとテキストIDとともに要素テキストリスト4cに登録する。ステップS10及びステップS11の処理により、要素テキストは、重複することなく要素テキストリスト4cに登録される。

【0063】

続いて、部分構造抽出モジュール15cは、処理対象要素を親要素として有する部分構造を構造化文書内部表現13から抽出し、その部分構造の部分構造表現を取得する（ステップS18）。既述の通り、部分構造表現は、要素IDと属性IDで表現されている。部分構造リスト4eの作成に使用される要素IDは、要素名を検索キーとして要素名リスト4aを検索することによって取得され、属性IDは、属性名を検索キーとして属性名リスト4bを検索することによって取得される。

【0064】

更に部分構造抽出モジュール15cは、抽出した部分構造が部分構造リスト4eに登録されているかを調べる（ステップS19）。登録されていない場合、部分構造抽出モジュール15cは、該部分構造に構造IDを与え、更に、該部分構造の部分構造表現を、構造IDとともに部分構造リスト4eに登録する。

【0065】

続いて、構造化文書符号化モジュール12は、処理対象要素が子要素を有するかを調べる（ステップS21）。有する場合、構造化文書符号化モジュール12は、該処理対象要素の子要素のうちの一つを選択する（ステップS22）。更に、構造化文書符号化モジュール12は、選択された子要素を処理対象要素としてステップS02以下の手順を再帰的に実行する（ステップS23）。これにより、選択された子要素の要素名と要素IDが要素リスト4aに登録され、その子要素に与えられている属性の属性名と属性IDとが属性リスト4bに登録され、その属性の属性テキスト（属性値）が属性テキストリスト4dに登録され、その子要素の要素テキストが要素テキストリスト4cに登録され、その子要素を親要素とする部分構造が部分構造リスト4dに登録される。選択された子要素が、更に子要素を有する場合も同様の処理が行われる。

【0066】

全ての要素についてステップ S 0 2 からステップ S 2 3 の処理が行われると、全体文書構造抽出モジュール 1 6 は、構造化文書内部表現 1 3 を構造 I D とテキスト I D とで表現する全体構造データ 4 f を生成する。

【0067】

更に、符号化文書出力モジュール 1 7 が、要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、部分構造リスト 4 e 及び全体構造データ 4 f を統合して符号化文書 4 を生成し、符号化文書 4 の生成処理が完了する（ステップ S 2 6）。

【0068】

図 1 3 は、構造化文書復号化モジュール 2 2 が、符号化文書 4 から生成されるべき構造化文書内部表現 2 3 の妥当性を検証し、更に、符号化文書 4 から構造化文書内部表現 2 3 を復元する処理を示すフローチャートである。まず、構造化文書復号化モジュール 2 2 が符号化文書 4 を受け取ると、構造化文書復号化モジュール 2 2 の符号化文書入力モジュール 2 6 は、符号化文書 4 から、要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、部分構造リスト 4 e、及び全体構造データ 4 f を抽出する（ステップ S 3 2 ～ S 3 7）。

【0069】

更に、構造化文書復号化モジュール 2 2 のテキスト型検証モジュール 2 5 a は、要素テキストリスト 4 c に列挙された要素テキストを検証する（ステップ S 3 8）。要素テキストの検証は、下記の手順で行われる。テキスト型検証モジュール 2 5 a は、要素テキストリスト 4 c に列挙された要素テキストを有する要素の要素名を、要素 I D を検索キーとして要素名リスト 4 a を検索して取得する。更にテキスト型検証モジュール 2 5 a は、取得した要素名を検索キーとして文書構造定義ファイル 2 4 を検索し、その要素名を有する要素の要素型宣言を取得する。更にテキスト型検証モジュール 2 5 a は、要素テキストが、要素型宣言に規定された型に適合しているかを判定する（ステップ S 3 9）。要素テキストが、要素型宣言に規定された型に適合していない場合、即ち、要素テキストに要素型違反がある場合、要素型違反の発生を通知する要素型違反通知を生成する。要素型

違反通知は、受信側コンピュータ 2 の表示装置に表示され、更に、送信側コンピュータ 1 に送信されて送信側コンピュータ 1 の表示装置に表示される（ステップ S 4 4）。要素型違反が発生した場合、構造化文書内部表現 2 3 の復元は行われず、構造化文書復号化モジュール 2 2 の処理は終了する（ステップ S 4 7）。

【0070】

要素型違反が発生していない場合、テキスト型検証モジュール 2 5 a は、更に、属性テキストリスト 4 d に列挙された属性テキストを検証する（ステップ S 4 0）。属性テキストの検証は下記の手順で行われる。テキスト型検証モジュール 2 5 a は、属性テキストリスト 4 d に列挙された属性テキストを属性値として有する属性の属性名を、属性 ID を検索キーとして属性名リスト 4 b を検索して取得する。更にテキスト型検証モジュール 2 5 a は、取得した属性名を検索キーとして文書構造定義ファイル 2 4 を検索し、その属性名を有する属性の属性型宣言を取得する。更にテキスト型検証モジュール 2 5 a は、属性テキストが、属性型宣言に規定された型に適合しているかを判定する（ステップ S 4 1）。属性テキストが、属性型宣言に規定された型に適合していない場合、即ち、属性テキストに属性型違反がある場合、属性型違反の発生を通知する属性型違反通知を生成する。属性型違反通知は、受信側コンピュータ 2 の表示装置に表示され、更に、送信側コンピュータ 1 に送信されて送信側コンピュータ 1 の表示装置に表示される（ステップ S 4 4）。属性型違反が発生した場合、構造化文書内部表現 2 3 の復元は行われず、構造化文書復号化モジュール 2 2 の処理は終了する（ステップ S 4 7）。

【0071】

属性型違反が発生していない場合、部分構造検証モジュール 2 5 b は、部分構造リスト 4 e に列挙されている部分構造を検証する（ステップ S 4 2）。部分構造の検証は、下記の手順で行われる。部分構造検証モジュール 2 5 b は、部分構造リスト 4 e に列挙された部分構造の部分構造表現に現れている要素 ID を検索キーとして要素名リスト 4 a を検索する。部分構造検証モジュール 2 5 b は、要素名リスト 4 a を検索することにより、その部分構造に関係している要素の要素名を取得する。更に、その部分構造に関係している要素のうち、親要素の要素名

を検索キーとして文書構造定義ファイル 24 を検索し、その部分構造が取るべき型を取得する。部分構造検証モジュール 25 b は、その部分構造が、文書構造定義ファイル 24 に規定された型に適合しているかを判定する（ステップ S 4 3）。部分構造が、文書構造定義ファイル 24 に規定されている型に適合しない場合、部分構造の型違反の発生を通知する部分構造型違反通知を生成する。部分構造型違反通知は、受信側コンピュータ 2 の表示装置に表示され、更に、送信側コンピュータ 1 に送信されて送信側コンピュータ 1 の表示装置に表示される（ステップ S 4 4）。部分構造型違反が発生した場合、構造化文書内部表現 23 の復元は行われず、構造化文書復号化モジュール 22 の処理は終了する（ステップ S 4 7）。

【0072】

要素型違反、属性型違反、及び部分構造型違反のいずれも発生しなかった場合、構造化文書内部表現出力モジュール 27 は、要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、部分構造リスト 4 e 及び全体構造データ 4 f から構造化文書内部表現 23 を再生する。構造化文書内部表現出力モジュール 27 は、要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e を参照して全体構造データ 4 f を解析する。構造化文書内部表現出力モジュール 27 は、全体構造データ 4 f を解析することによって構造化文書内部表現 23 に含まれるべき要素の要素名、その要素の要素内容（即ち、要素テキスト及び子要素の要素名）、その要素に定められた属性の属性名、属性の属性テキスト（属性値）を把握し、更に構造化文書内部表現 23 を生成する。

【0073】

更に構造化文書内部表現出力モジュール 27 は、型違反が発生しなかったことを通知する正常通知を生成する。正常通知は、受信側コンピュータ 2 の表示装置に表示され、更に、送信側コンピュータ 1 に送信されて送信側コンピュータ 1 の表示装置に表示される（ステップ S 4 6）。構造化文書復号化モジュール 22 の処理は終了する（ステップ S 4 7）。

【0074】

このように、構造化文書復号化モジュール 2 2 は、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e に基づいて構造化文書内部表現 2 3 の妥当性を検証し、その妥当性が保証された後に構造化文書内部表現 2 3 を復元する。このような検証方法は、構造化文書内部表現 2 3 の妥当性を保証するために必要な演算の演算量を有効に減少する。上述のように、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e からは、重複が排除されている。要素テキストリスト 4 c において、同一の要素テキストが重複して列挙されないことは、該要素テキストの妥当性の検証を重複して行う必要をなくす。更に、属性テキストリスト 4 d において、同一の属性テキストが重複して列挙されないことは、該属性テキストの妥当性の検証を重複して行う必要をなくす。加えて、部分構造リスト 4 e において、同一の部分構造が重複して列挙されないことは、該部分構造の妥当性の検証を重複して行う必要をなくす。このように、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e における重複の排除により、構造化文書内部表現 2 3 の妥当性を保証するために必要な演算の演算量が抑制される。

【 0 0 7 5 】

以下では、構造化文書の処理の実施例が、説明される。

【 0 0 7 6 】

【実施例】

本実施例では、図 1 4 に示されている XML 文書について上記の実施の形態が適用される。該 XML 文書の構造がアプリケーションプログラム 1 1 によって解析され、構造化文書内部表現 1 3 が生成される。生成された構造化文書内部表現は、構造化文書符号化モジュール 1 2 に引き渡される

【 0 0 7 7 】

ステップ S 0 1 において要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e が初期化され、更に、ルート要素 " A d d r e s s b o o k " が処理対象要素として選択される。ステップ S 0 2 においてルート要素の要素名 " A d d r e s s b o o k " が抽出される。この要素名は要素名リスト 4 a に登録されていない。従って、ステッ

プ S 0 4 において、図 1 5 に示されているように、ルート要素の要素 ID " E 0 1 " と要素名 " A d d r e s s b o o k " とが要素名リスト 4 a に登録される。

【 0 0 7 8 】

ステップ S 0 5 において処理対象要素であるルート要素が属性を有するかが調べられる。図 1 4 に示されているように、ルート要素は、属性 " o w n e r " を有する。ステップ S 0 6 において属性 " o w n e r " が選択され、ステップ S 0 7 においてその属性の属性名 " o w n e r " が抽出される。その属性名は、属性リスト 4 b に登録されていない（ステップ S 0 8）。そのため、ステップ S 0 9 において、図 1 6 に示されているように、該属性の属性 ID " A 0 1 " と属性名 " o w n e r " とが属性名リスト 4 b に登録される。更に、ステップ S 1 0 において、該属性の属性テキスト " S u z u k i " が抽出される。属性テキスト " S u z u k i " は、属性テキストリスト 4 d に登録されていない（ステップ S 1 1）。そのため、ステップ S 1 2 において、図 1 8 に示されているように、該属性の属性 ID " A 0 1 " と、該属性テキストのテキスト ID " T 1 0 " と、属性テキスト " S u z u k i " とが属性テキストリスト 4 d に登録される。ルート要素は、他の属性を有しないため（ステップ S 1 3）、ステップ S 0 6 からステップ S 1 2 までの処理は完了する。

【 0 0 7 9 】

続いて、ステップ S 1 4 において、処理対象要素であるルート要素が要素テキストを有するかが判断される。ルート要素は、その要素内容として要素テキストを有しないので、処理がステップ S 1 8 にジャンプする。

【 0 0 8 0 】

ステップ S 1 8 では、ルート要素 " A d d r e s s b o o k " を親要素として有する部分構造が抽出される。即ち、ルート要素に与えられた属性と、ルート要素の子要素とが抽出される。該部分構造の部分構造表現は親要素（ルート要素）の要素 ID、親要素の属性の属性 ID、及び子要素の要素 ID で記述される。ルート要素の子要素 " P e r s o n " の要素 ID を " E 0 2 " とすると、ルート要素を親要素として有する部分構造の部分構造表現は、例えば、 " E 0 1 A 0 1 E 0 2 E 0 2 " と定められる。本実施例では、部分構造表現は、親要素の要

素 ID、親要素が有する属性の属性 ID、及び子要素の要素 ID が、この順に記載されるという規則に従って作成されている。親要素が属性を複数有する場合には、その属性 ID が順次に並べられ、親要素が子要素を複数有する場合には、その要素 ID が順次に並べられる。

【0081】

更に、ステップ S19 では抽出された部分構造が部分構造リスト 4e に登録されているかが判断される。ルート要素” Addressbook” を親要素として有する部分構造は登録されていないので、図 19 に示されているように、ステップ S20 において該部分構造に構造 ID” S01” が与えられ、更に、該部分構造の部分構造表現” E01 A01 E02 E02” が、構造 ID” S01” とともに部分構造リスト 4e に登録される。

【0082】

続いて、ステップ S21 においてルート要素” Addressbook” が子要素を有するかが判断される。図 14 に示されているように、ルート要素は、子要素” Person” を 2 つ有している。子要素” Person” が処理対象要素として順次に選択され、ステップ S02 から S22 までの処理が、再帰的に実行される（ステップ S22～24）。

【0083】

図 14 の XML 文書を基にして生成される要素名リスト 4a が図 15 に示され、属性名リスト 4b が図 16 に示され、要素テキストリスト 4c が図 17 に示され、属性テキストリスト 4d が図 18 に示され、部分構造リスト 4e が図 19 に示されている。

【0084】

更に、全体構造データ 4f が生成される。図 14 に示されている XML 文書の全体構成は、図 19 の部分構造リスト 4e に示されている構造 ID と、図 17、図 18 の要素テキストリスト 4c 及び属性テキストリスト 4d に列挙されているテキスト ID とを用いて、例えば、

” S01 T10 S02 S03 S04 T01 S05 T02 S06
S07 T04 S08 T05 S09 T06 S10 T08 S11

S 0 3 S 0 4 T 0 1 S 0 5 T 0 3 S 1 2 S 0 7 T 0 4 S 0 8
T 0 5 S 0 9 T 0 7 S 1 3 T 0 9”

と表現される。本実施例では、下記の規則に従って全体構造データ 4 f が生成されている。最上位の階層の要素を親要素とする部分構造の構造 ID が先頭に置かれ、該親要素の属性の属性テキスト及び該親要素の要素テキストのテキスト ID が次に並べられる。更に、該親要素の子要素を親要素とする部分構造の構造 ID が並べられる。以下再帰的に、該子要素の属性の属性テキスト及び該子要素の要素テキストのテキスト ID が並べられ、該子要素の子要素を親要素とする部分構造の構造 ID が並べられる。

【0085】

上記の全体構造データ 4 f を理解しやすくするために括弧を挿入すると、全体構造データ 4 f は、” S 0 1 T 1 0 (S 0 2 (S 0 3 (S 0 4 T 0 1
) (S 0 5 T 0 2)) (S 0 6 (S 0 7 T 0 4) (S 0 8 T 0 5
) (S 0 9 T 0 6)) (S 1 0 T 0 8)) (S 1 1 (S 0 3 (S
0 4 T 0 1) (S 0 5 T 0 3)) (S 1 2 (S 0 7 T 0 4) (S
0 8 T 0 5) (S 0 9 T 0 7) (S 1 3 T 0 9)))) ” と表現される。

【0086】

先頭の” S 0 1 ” は、構造 ID ” S 0 1 ” を有する部分構造に含まれる親要素がルート要素であることを示している。構造 ID ” S 0 1 ” が与えられている部分構造の部分構造表現” E 0 1 A 0 2 E 0 2 E 0 2 ” は、要素 ID ” E 0 1 ” を有する要素（要素” A d d r e s s b o o k ”、図 15 参照）がルート要素であることを示している。ルート要素の要素名は、図 15 に示されている要素名リスト 4 a から” A d d r e s s b o o k ”であることがわかる。

【0087】

更に、構造 ID ” S 0 1 ” を有する部分構造の部分構造表現” E 0 1 A 0 2
E 0 2 E 0 2 ” は、そのルート要素は、属性 ID ” A 0 1 ” を有する属性（属性” o w n e r ”、図 16 参照）を有し、且つ、要素 ID ” E 0 2 ” を有する要素（要素” P e r s o n ”、図 15 参照）を 2 つ有していることを示している

。全体構造データ 4 f の” S 0 1 ” に続く” T 1 0 ” は、その属性の属性値がテキスト ID ” T 1 0 ” を有する属性テキスト” S u z u k i ”（図 1 8 参照）であることを示している。

【0088】

要素 ID ” E 0 2 ” を有する 2 つの要素” P e r s o n ” のうちの最初の一方には、表現（S 0 2 …）が割り当てられ、2 番目の方には、表現（S 1 1 …）が割り当てられる。表現（S 0 2 …）に含まれる構造 ID ” S 0 2 ” を有する部分構造の部分構造表現” E 0 2 E 0 3 E 0 6 E 1 0 ” から、要素” P e r s o n ” の最初の一方は、要素 ID ” E 0 3 ”、” E 0 6 ”、及び” E 1 0 ” をそれぞれに有する要素を子要素として有していることが分かる。要素 ID ” E 0 3 ” を有する要素には、表現（S 0 3 …）が割り当てられ、要素 ID ” E 0 6 ” を有する要素には、表現（S 0 6 …）が割り当てられ、要素 ID ” E 1 0 ” を有する要素には、表現（S 1 0 …）が割り当てられる。以下、同様にして図 1 4 の XML 文書の全体構造が表現される。但し、このような全体構造を表現するために、他の規則が使用され得ると理解されなくてはならない。

【0089】

図 1 5 乃至図 1 9 に示されている要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e と、上記の全体構造データ 4 f とが統合されて、符号化文書 4 が生成される。

【0090】

一方、図 1 4 の XML 文書に対応する符号化文書 4 が構造化文書復号化モジュール 2 2 に入力されると、ステップ S 3 2 ～ S 3 7 において、その符号化文書 4 から、図 1 5 乃至図 1 9 に示されている要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e と、上記の全体構造データ 4 f とが取り出される。

【0091】

続いて、ステップ S 3 8 において、テキスト型検証モジュール 2 5 a により、要素テキストリスト 4 c に挙げられている要素テキストの型が検証される。要素

テキストリスト 4 c に挙げられている 9 個の要素テキストのそれぞれについて、その要素テキストを要素内容とする要素の要素 ID が取り出され、取り出された要素 ID を検索キーとして要素名リスト 4 a が検索されて、該要素の要素名が取り出される。その要素名を用いて文書構造定義ファイル 2 4 が検索されて、該要素名を有する要素の要素型宣言が取り出される。更に、要素テキストが該要素型宣言に規定されている型に適合するかが検証される。

【0092】

例えば、図 1 5 を参照して、要素” L a s t N a m e ” について考える。更に、文書構造定義ファイル 2 4 に規定された要素” L a s t N a m e ” の要素型宣言により、要素” L a s t N a m e ” の要素テキストが文字型であると規定されているとする。図 1 7 を参照すると、要素” L a s t N a m e ” は要素テキスト” Y a m a d a ” を有しており、要素” L a s t N a m e ” の要素テキストは、正しい型を有していると判断される。

【0093】

更に例えば、図 1 5 を参照して、要素” T e l e p h o n e ” について考える。文書構造定義ファイル 2 4 に規定された要素” T e l e p h o n e ” の要素型宣言により、要素” T e l e p h o n e ” の要素テキストが、3 桁の整数と、それに続くハイフン” - ” と、それに続く 3 桁の整数と、それに続くハイフンと、それに続く 4 桁の整数というフォーマットを有すると規定されているとする。図 1 7 を参照すると、要素” T e l e p h o n e ” は要素テキスト” 0 4 5 - 1 2 3 - 4 5 6 7 ” を有しており、要素” T e l e p h o n e ” の要素テキストは、正しい型を有していると判断される。

【0094】

要素テキストが取り得る型は、この他にも、例えば、整数型、浮動小数点型、日付型であり得る。

【0095】

同様に、全ての要素について要素型違反がないと判断されると、ステップ S 4 0 において、属性テキストリスト 4 d に挙げられている属性テキストの型が検証される。属性テキストリスト 4 d に挙げられている属性テキストのそれぞれにつ

いて、その属性テキストを属性値とする属性の属性 ID が取り出され、取り出された属性 ID を検索キーとして属性名リスト 4 b が検索されて、該属性の属性名が取り出される。その属性名を用いて文書構造定義ファイル 2 4 が検索されて、該属性名を有する属性の属性型宣言が取り出される。更に、属性テキストが該属性型宣言に規定されている型に適合するかが検証される。

【 0 0 9 6 】

例えば、図 1 6 を参照して、属性 " O w n e r " について考える。更に、文書構造定義ファイル 2 4 に規定された属性 " O w n e r " の属性型宣言により、属性 " O w n e r " の属性テキスト（属性値）が文字型であると規定されているとする。図 1 8 を参照すると、属性 " O w n e r " の要素テキストは、" S u z u k i " であり、属性 " O w n e r " の要素テキストは、正しい型を有していると判断される。

【 0 0 9 7 】

更に、ステップ S 4 2 において部分構造リスト 4 e に列挙されている 1 3 個の部分構造が検証される。例えば、図 1 9 を参照して、構造 ID " S 0 1 " を有する部分構造の部分構造表現は、" E 0 1 A 0 1 E 0 2 E 0 2 " である。要素名リスト 4 a と属性名リスト 4 b とを参照すると、この部分構造表現は、要素 ID " E 0 1 " を有する要素 " A d d r e s s b o o k " は属性 " O w n e r " を有し、且つ、その子要素が 2 つの要素 " P e r s o n " であることを示している。この部分構造が、文書構造定義ファイル 2 4 に規定されている構文に適合するかが判定される。他の部分構造についても、同様の検証が行われる。

【 0 0 9 8 】

以上の工程により、要素テキスト、属性テキスト、及び部分構造の検証が終了する。検証が終了した後、要素名リスト 4 a、属性名リスト 4 b、要素テキストリスト 4 c、属性テキストリスト 4 d、及び部分構造リスト 4 e と、上記の全体構造データ 4 f から、図 1 4 の XML 文書の構造化文書内部表現 2 3 が復元される。

【 0 0 9 9 】

このような検証方法は、同一の要素テキスト、属性テキスト、及び部分構造に

対して重複して検証することを避けることができる。図14のXML文書を用いてその妥当性を検証しようとするれば、同一の検証対象に対して重複した検証が行われる。例えば、要素"Person"が"要素"Name"と要素"Address"を有し、且つ、要素"Name"が要素"LastName"と要素"FirstName"を有するという記述は、図14のXML文書にそれぞれ2回現れる。従って、図14のXML文書を用いて検証が行われた場合、これらの記述は、2回検証されたはずである。更に、要素"LastName"の要素テキスト"Yamada"の型が正しいかという検証も2回繰り返されたはずである。要素"Country"の要素テキスト"Japan"、要素"Prefecture"の要素テキスト"Kanagawa"についても同様である。

【0100】

本実施の形態では、冗長性を排除して構造化文書を表現した符号化文書4が送信され、その符号化文書4に基づいて再生されるべき構造化文書内部表現23の妥当性が検証され、妥当性が保証された後に構造化文書内部表現23が再生される。これにより、同一の検証作業の繰り返しが排除され、検証作業のコストを低減することができる。

【0101】

図20に示されているように、本実施の形態のコンピュータシステムは、送信側コンピュータ1にデータ圧縮モジュール18を有し、受信側コンピュータ2にデータ解凍モジュール28を有していることが好適である。データ圧縮モジュール18は、符号化文書4を圧縮して圧縮符号化文書4'を生成する。データ解凍モジュール28は、圧縮符号化文書4'を受信して解凍し、符号化文書4を再生する。このような構成は、送信側コンピュータ1から受信側コンピュータ2に通信されるデータの量を抑制する。

【0102】

本実施の形態のコンピュータシステムにおいて、要素IDが一連番号を含んで構成される場合、要素名リスト4aには明示的に要素IDが記述される必要はない。要素IDの代わりに、要素名リスト4a内において要素名が記述される順番を使用することが可能である。例えば、要素名リスト4aにおいて第1番目に記

述される要素名は、要素 I D” E 0 1” を有すると規定され、その要素 I D が要素名リスト 4 a に記述されないことが可能である。通信コストに敏感な状況において本実施の形態のコンピュータシステムが使用される場合、要素名リスト 4 a をより小さなデータ量で表現するメリットは大きい。

【0103】

同様に、属性 I D が一連番号を含んで構成される場合、属性名リスト 4 b には、明示的に要素 I D が記述される必要はない。

【0104】

更に同様に、要素テキストリスト 4 c 及び属性テキストリスト 4 d が一のリストに記述され、且つ、テキスト I D が一連番号を含んで構成される場合、要素テキストリスト 4 c 及び属性テキストリスト 4 d には、明示的にテキスト I D が記述される必要はない。

【0105】

更に同様に、構造 I D が一連番号を含んで構成される場合、部分構造リスト 4 e には、構造 I D が明示的に記述される必要がない。

【0106】

【発明の効果】

本発明により、構造化文書を検証するために必要な演算量を抑制し、これにより、検証に必要なコストを減少することができる。

また、本発明により、構造化文書を送信側コンピュータから受信側コンピュータに伝送する場合に、受信側コンピュータが構造化文書を検証するために必要な演算量を抑制し、これにより、検証に必要なコストを減少することができる。

【図面の簡単な説明】

【図1】

図1は、本発明によるコンピュータシステムの実施の一形態を示すブロック図である。

【図2】

図2は、符号化文書4の内容を示す。

【図3】

図 3 は、要素名リスト 4 a の内容を示す。

【図 4】

図 4 は、属性名リスト 4 b の内容を示す。

【図 5】

図 5 は、要素テキストリスト 4 c の内容を示す。

【図 6】

図 6 は、属性テキストリスト 4 d の内容を示す。

【図 7】

図 7 は、部分構造リスト 4 e の内容を示す。

【図 8】

図 8 は、構造化文書符号化モジュール 1 2 の構成を示すブロック図である。

【図 9】

図 9 は、構造化文書復号化モジュール 2 2 の構成を示すブロック図である。

【図 10】

図 10 は、本実施の形態のコンピュータシステムの変形例を示すブロック図である。

【図 11】

図 11 は、構造化文書内部表現 1 3 から符号化文書 4 を生成する手順を示すフローチャートである。

【図 12】

図 12 は、構造化文書内部表現 1 3 から符号化文書 4 を生成する手順を示すフローチャートである。

【図 13】

図 13 は、符号化文書 4 に基づいて構造化文書内部表現 2 3 の妥当性を検証し、更に、構造化文書内部表現 2 3 を再構成する手順を示すフローチャートである。

【図 14】

図 14 は、XML 文書の例を示す。

【図 15】

図 1 5 は、図 1 4 の XML 文書の内部表現から生成される要素名リスト 4 a を示す。

【図 1 6】

図 1 6 は、図 1 4 の XML 文書の内部表現から生成される属性名リスト 4 b を示す。

【図 1 7】

図 1 7 は、図 1 4 の XML 文書の内部表現から生成される要素テキストリスト 4 c を示す。

【図 1 8】

図 1 8 は、図 1 4 の XML 文書の内部表現から生成される属性テキストリスト 4 d を示す。

【図 1 9】

図 1 9 は、図 1 4 の XML 文書の内部表現から生成される部分構成リスト 4 e を示す。

【図 2 0】

図 2 0 は、本実施の形態のコンピュータシステムの他の変形例を示すブロック図である。

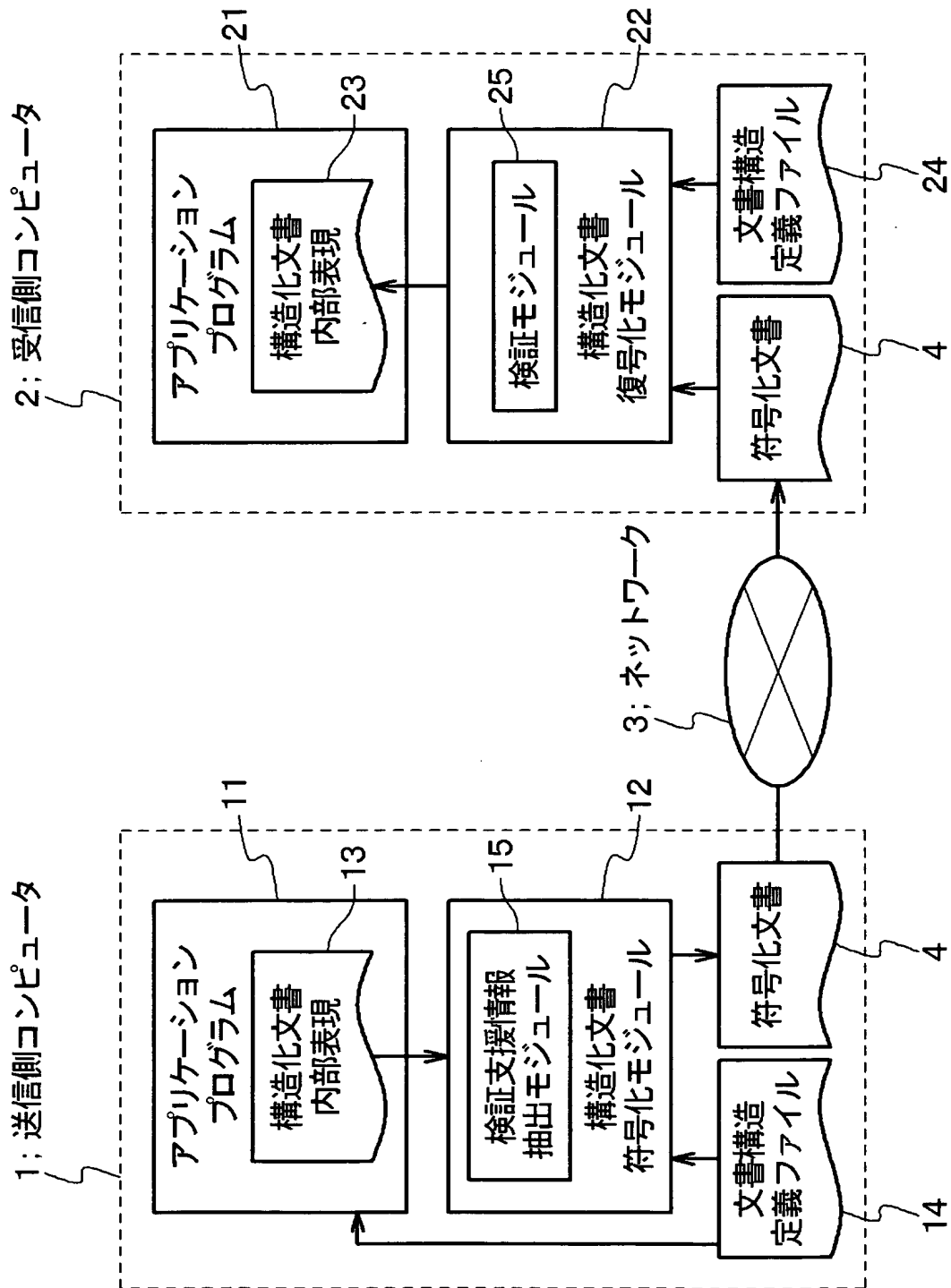
【符号の説明】

- 1：送信側コンピュータ
- 2：受信側コンピュータ
- 3：ネットワーク
- 4：符号化文書
- 4 a：要素名リスト
- 4 b：属性名リスト
- 4 c：要素テキストリスト
- 4 d：属性テキストリスト
- 4 e：部分構造リスト
- 4 f：全体構造データ
- 11：アプリケーションプログラム

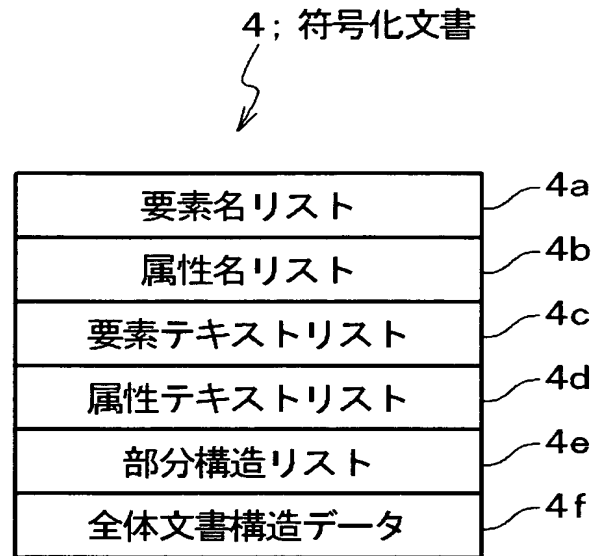
- 1 2 : 構造化文書符号化モジュール
- 1 3 : 構造化文書内部表現
- 1 3' : 構造化文書
- 1 4 : 文書構造定義ファイル
- 1 5 : 検証支援情報抽出モジュール
- 1 5 a : 要素名／属性名抽出モジュール
- 1 5 b : テキスト抽出モジュール
- 1 5 c : 部分構造抽出モジュール
- 1 6 : 全体文書構造抽出モジュール
- 1 7 : 符号化文書出力モジュール
- 2 1 : アプリケーションプログラム
- 2 2 : 構造化文書復号化モジュール
- 2 3 : 構造化文書内部表現
- 2 4 : 文書構造定義ファイル
- 2 5 : 検証モジュール
- 2 5 a : テキスト型検証モジュール
- 2 5 b : 部分構造検証モジュール
- 2 6 : 符号化文書入力モジュール
- 2 7 : 構造化文書内部表現出力モジュール

【書類名】 図面

【図 1】



【図 2】



【図 3】

4a; 要素名リスト

The table has two columns: '要素 ID' (Element ID) and '要素名' (Element Name). It lists four elements: E01 (要素名 A), E02 (要素名 B), E03 (要素名 C), and E04 (要素名 D). The last row contains three dots in both columns. Above the table, the text '4a; 要素名リスト' (4a; Element Name List) is written, with a curved arrow pointing down to the table.

要素 ID	要素名
E01	要素名 A
E02	要素名 B
E03	要素名 C
E04	要素名 D
...	...

【図 4】

4b; 属性名リスト



属性 I D	属性名
A01	属性名 A
A02	属性名 B
A03	属性名 C
A04	属性名 D
.

【図 5】

4c; 要素テキストリスト



要素 I D	テキスト I D	要素テキスト
E01	T01	テキスト A
E01	T02	テキスト B
E01	T03	テキスト C
E02	T04	テキスト D
.

【図 6】

4d; 属性テキストリスト



属性 I D	テキスト I D	属性テキスト
A01	T05	テキスト A
A01	T06	テキスト B
A01	T07	テキスト C
A02	T08	テキスト D
.

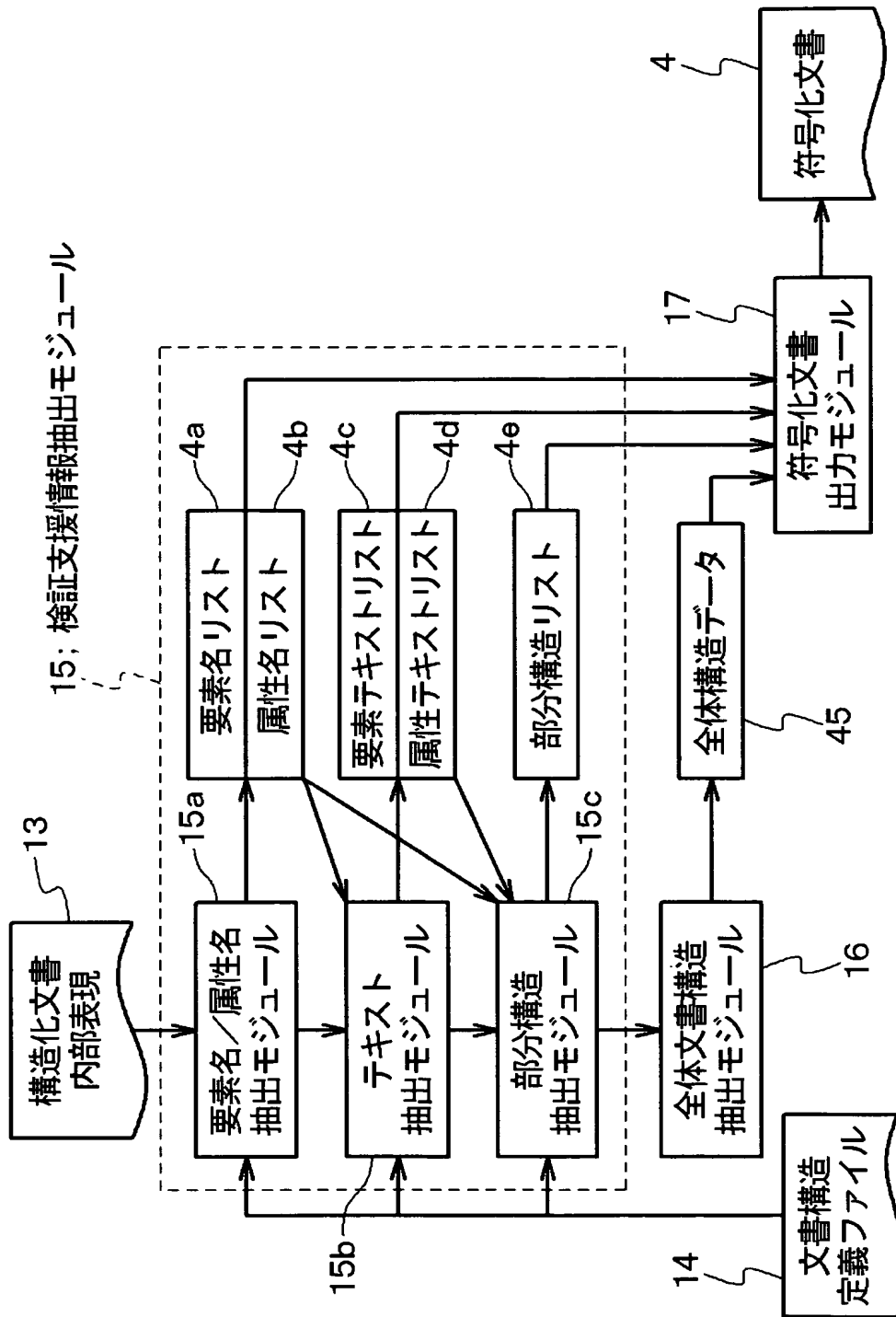
【図 7】

4e; 部分構造リスト

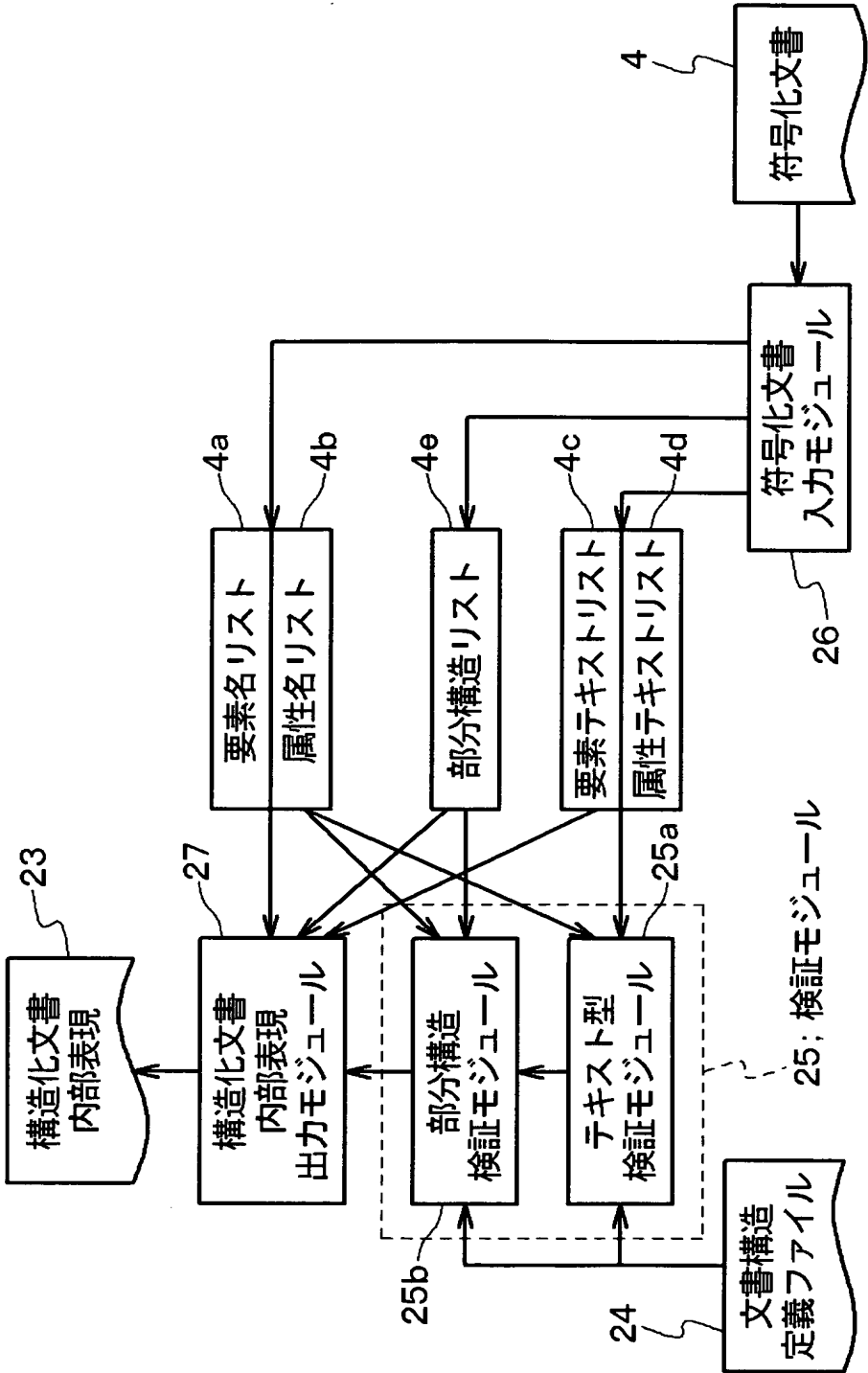


構造 I D	部分構造表現
S01	E01 A01 E02 E02
S02	E02 E03 E06 E10
S03	E03 E04 E05
S04	E04 T
.

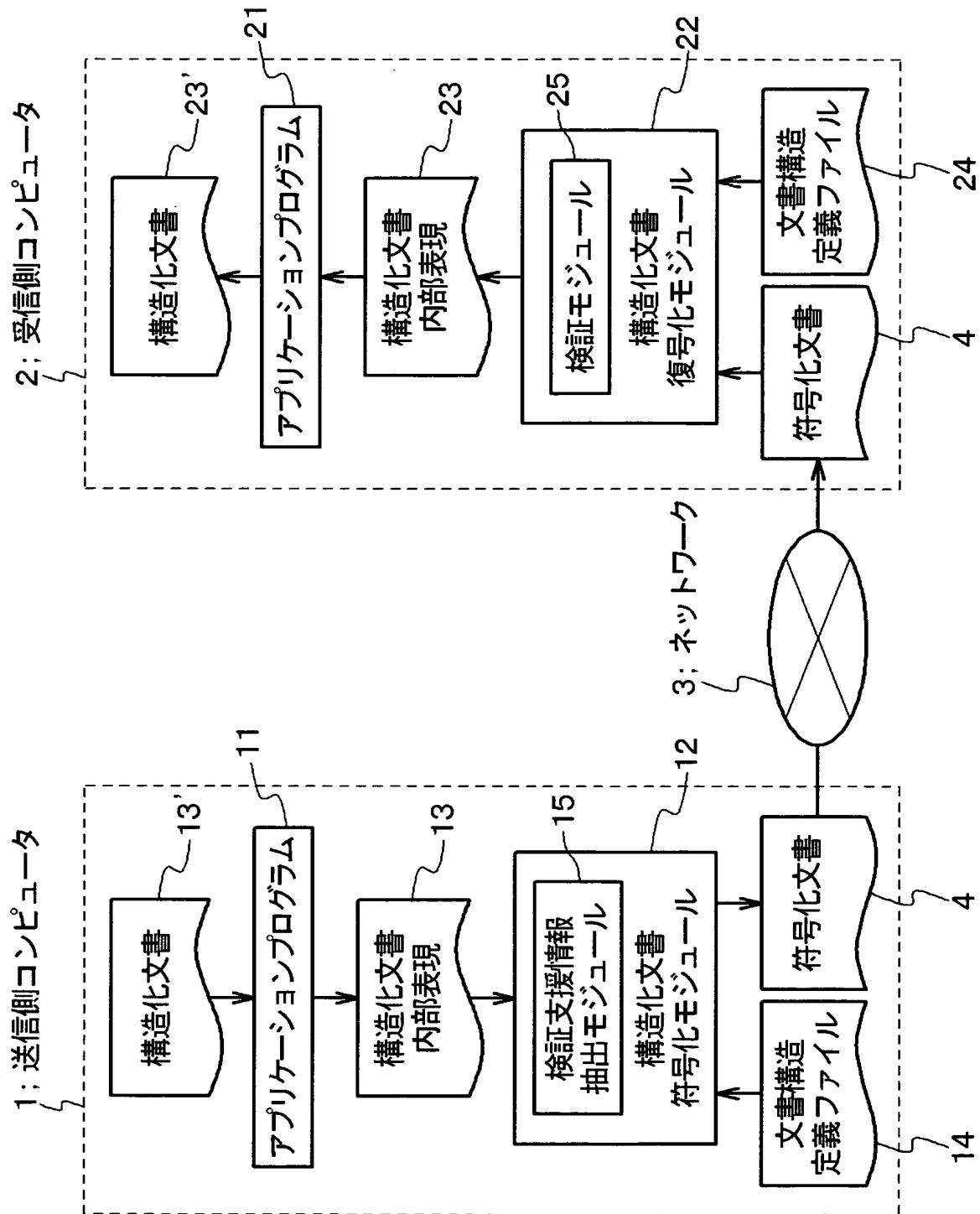
【図 8】



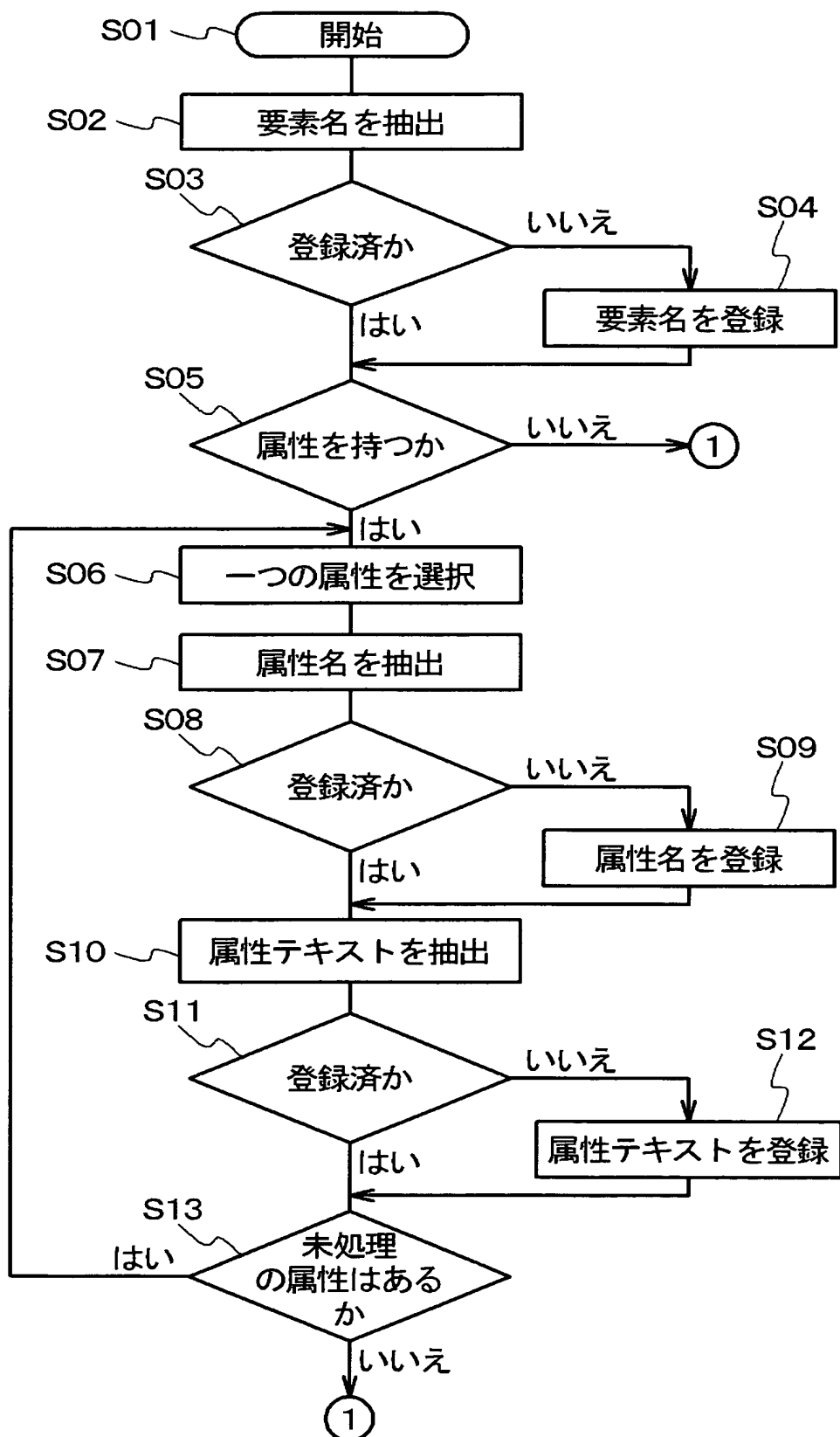
【図 9】



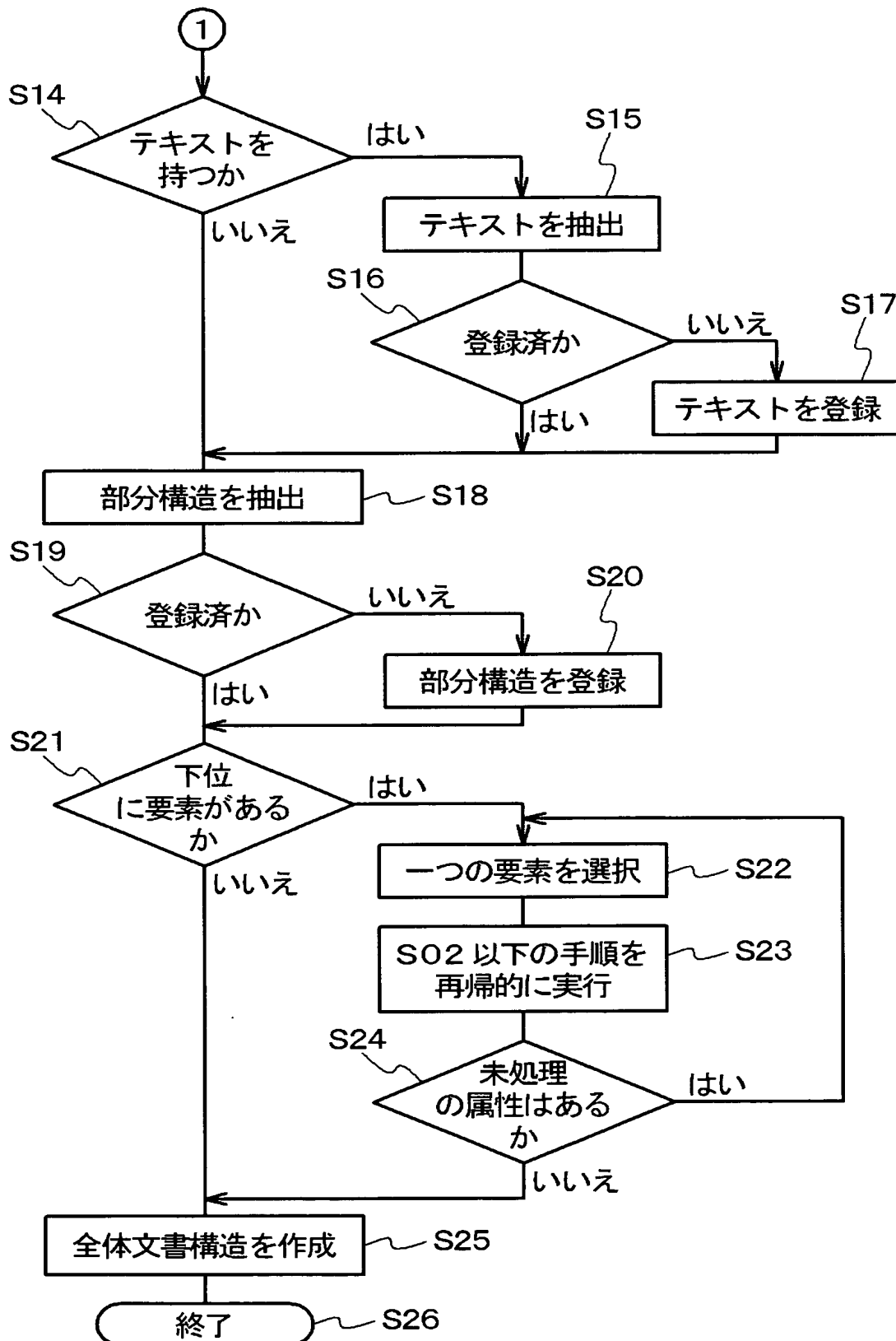
【図 10】



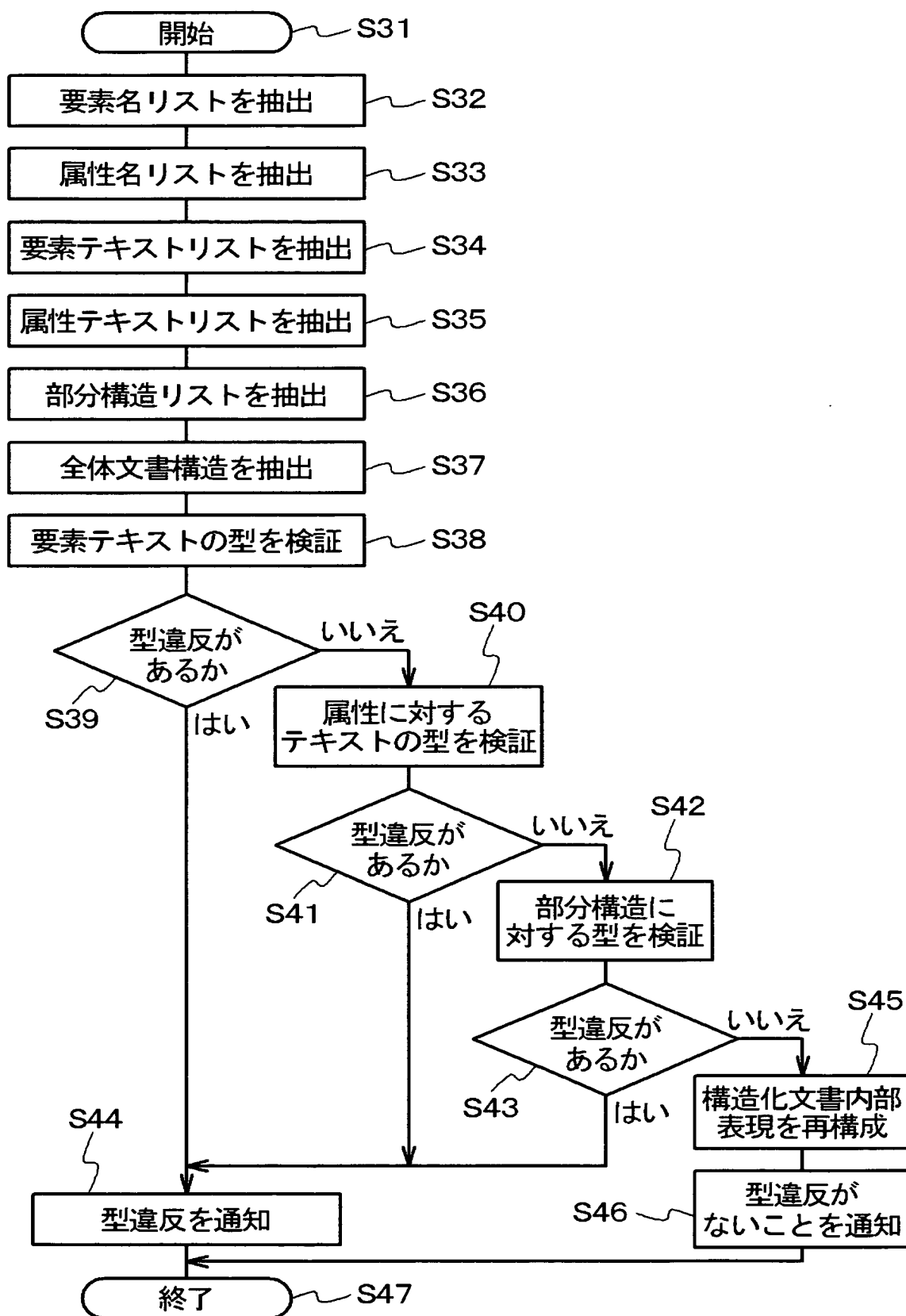
【図 11】



【図 12】



【図 13】



【図 1 4】

```
<AddressBook owner = "Suzuki">
  <Person>
    <Name>
      <LastName>Yamada</LastName>
      <FirstName>Taro</FirstName>
    </Name>
    <Address>
      <Country>Japan</Country>
      <Prefecture>Kanagawa</Prefecture>
      <City>Yokohama</City>
    </Address>
    <Telephone>045-123-4567</Telephone>
  </Person>
  <Person>
    <Name>
      <LastName>Yamada</LastName>
      <FirstName>Jiro</FirstName>
    </Name>
    <Address>
      <Country>Japan</Country>
      <Prefecture>Kanagawa</Prefecture>
      <City>Kawasaki</City>
      <Ward>Miyamae</Ward>
    </Address>
  </Person>
</AddressBook>
```

【図 1 5】

要素 I D	要素名
E01	AddressBook
E02	Person
E03	Name
E04	LastName
E05	FirstName
E06	Address
E07	Country
E08	Prefecture
E09	City
E10	Telephone
E11	Ward

【図 1 6】

属性 I D	属性名
A01	Owner

【図 17】

要素ID	テキストID	要素テキスト
E04	T01	Yamada
E05	T02	Taro
E05	T03	Jiro
E07	T04	Japan
E08	T05	Kanagawa
E09	T06	Yokohama
E09	T07	Kawasaki
E10	T08	045-123-4567
E11	T09	Miyamae

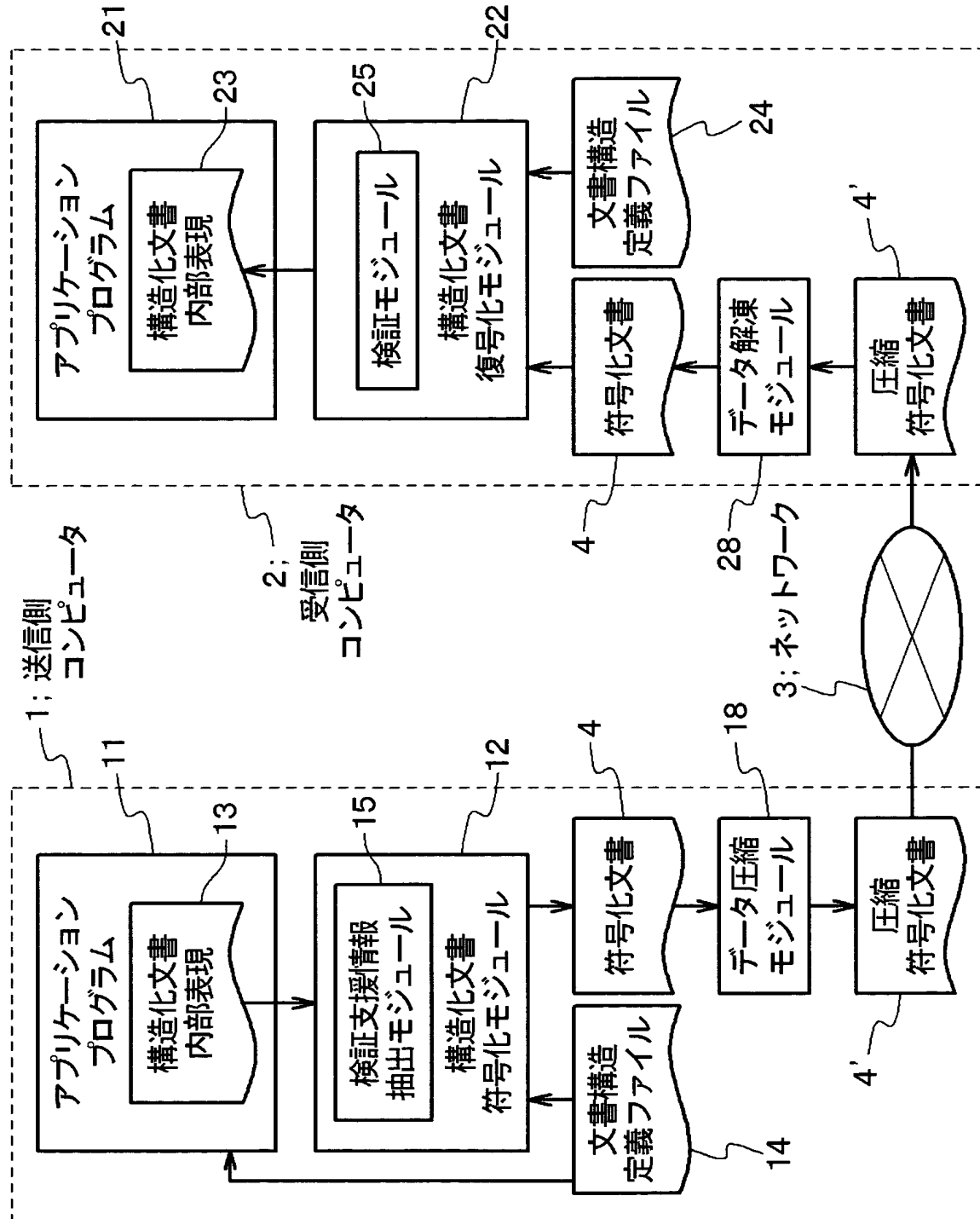
【図 18】

属性ID	テキストID	属性テキスト
A01	T10	Suzuki

【図 1 9】

構造 I D	部分構造表現
S01	E01 A01 E02 E02
S02	E02 E03 E06 E10
S03	E03 E04 E05
S04	E04 T
S05	E05 T
S06	E06 E07 E08 E09
S07	E07 T
S08	E08 T
S09	E09 T
S10	E010 T
S11	E02 E03 E06
S12	E06 E07 E08 E09 E11
S13	E11 T

【図 20】



【書類名】 要約書

【要約】

【課題】 ネットワークを介して通信された構造化文書を検証するために必要な演算量を抑制する。

【解決手段】 本発明によるコンピュータシステムは、構造化文書の内部表現（13）を符号化して符号化文書（4）を作成する符号化手段（12）を含む送信側コンピュータ（1）と、ネットワーク（3）を介して符号化文書（4）を取得し、符号化文書（4）を復号して内部表現（23）を再構成する復号手段（22）を含む受信側コンピュータ（2）とを備えている。符号化手段（12）は、該構造化文書に含まれる記述を重複することなく列挙することによって内部表現（13）から検証支援情報（4a～4e）を生成する検証支援情報抽出手段（15）を含み、且つ、検証支援情報（4a～4e）を含むように符号化文書（4）を生成する。復号手段（22）は、検証支援情報（4a～4e）に基づいて、再構成される内部表現（23）のデータ構造が、所定の文書型宣言に適合しているかを検証する検証手段（25）を含む。当該コンピュータシステムでは、記述の重複が排除された検証支援情報（4a～4e）に基づいて内部表現（23）のデータ構造の妥当性が検証される。このような構成は、検証手順の重複を排除し、従って、構造化文書を検証するために必要な演算量を抑制することができる。

【選択図】 図1

特願 2 0 0 3 - 0 9 4 2 8 1

出 願 人 履 歷 情 報

識別番号

[0 0 0 0 0 4 2 3 7]

1. 変更年月日

1 9 9 0 年 8 月 2 9 日

[変更理由]

新規登録

住 所

東京都港区芝五丁目 7 番 1 号

氏 名

日本電気株式会社